PROCEEDINGS

of

The 13th Annual Conference
MILITARY TESTING ASSOCIATION

Host

HEADQUARTERS
UNITED STATES MARINE CORPS
WASHINGTON, D. C. 20380

Statler Hilton Hotel
Washington, D. C.
20-24 September 1971

13th ANNUAL CONFERENCE OF THE MILITARY TESTING ASSOCIATION

0

20-24 September 1971

Statler Hilton Hotel
Washington, D. C.

P*R*O*G*R*A*M

MONDAY, 20 September 1971

0900-1700	Registration in Foyer 2
1300-1600	Steering Committee Meeting in the South American Room

TUESDAY, 21 September 1971

0800-0900	Registration in Foyer 2
0900-0920	Conference Called to Order in South American Room - Colonel L. T. ERICKSON, USMC, President, MTA
0920-0950	Opening Remarks by Commanders of MTA Organizations
0950-1015	Coffee Break in Foyer 2
1015-1045	Keynote Address - Lieutenant General O. R. SIMPSON, USMC, Deputy Chief of Staff (Manpower), U. S. Marine Corps
1045-1130	"Some Implications of the Supreme Court Decision, March 1971 (Civil Rights Act, 1964)" - Dr. RAYMOND O. WALDKOETTER, U.S. Army Enlisted Evaluation Center, Ft. Benjamin Harrison, Ind.
1300-1330	"USAF Officer Evaluation System--Review and Research Recommendation" - Major ROBERT E. WILKINSON, USAF, Air Force Human Resources Lab, Lackland AFB, Texas

TUESDAY, 21 September 1971 - continued

1335-1405	"Analysis of Enlisted Efficiency Report Trends" - Mr. KENNETH C. LIEBFRIED, U. S. Army Enlisted Evaluation Center, Ft. Benjamin Harrison, Ind.
1405-1450	"Automated Testing and Attrition Control II (ATAC II)" - Gunnery Sergeant DAVID A. DEORE, USMC, Marine Corps Communications-Electronics School, Twentynine Palms, California
1450-1515	Coffee Break
1515-1615	"The Use of Logic Trees in Military Performance Testing" - First Lieutenant RAYMOND L. ERICKSON, USA, U.S. Army Adjutant General School, Ft. Benjamin Harrison, Indiana
1615-1640	"Computer Simulation: A Tool for Psychologists" - Mr. WILLIAM A. SANDS, Naval Personnel Research and Development Laboratory, Washington, D.C.

WEDNESDAY, 22 September 1971

0900-0920	"U. S. Army Civilian Acquired Skills Testing Program" - Mr. JOHN BRAND, U.S. Army Enlisted Evaluation Center, Ft. Benjamin Harrison, Indiana
0920-0945	"Canadian Forces Personnel Selection Interview Study" - Major M. A. MARTIN, CAF, Personnel Applied Research Unit, Toronto, Ontario
0945-1015	"Clinical Evaluation and Prediction of Military Effectiveness of Naval Enlistees" - Commander (MC) ALFREDO BEYER R., Head, Pshcyological and Psychiatric Screening, Peruvian Navy
1015-1045	Coffee Break
1045-1120	"Systems Approach to Evaluation and Quality Control of Training" - LtCol BRYCE R. KRAMER, USA, U.S. Army Infantry School, Fort Benning, Georgia

WEDNESDAY, 22 September 1971 - continued

1120-1145	"General Training System (GENTRAS) Field Evaluation Routine" - Major JAMES K. MILLER, USMC, G-3 Division, Headquarters Marine Corps, Washington, D.C.
1300-1325	"The Development of the Navy Advisor Profile Report" - Mr. TED YELLEN, Naval Personnel Research and Development Laboratory, Washington, D.C.
1325-1340	"Factorial Profiles of the E8/E9 Examinations" - Mr. ERLING A. DUKERSCHEIN, U.S. Naval Examining Center, Great Lakes, Illinois
1340-1405	"Development of a Universal Equation for Predicting Job Difficulty" - Major DONALD F. MEAD, USAF, Air Training Command, Randolph Air Force Base, Texas
1405-1455	"Task Difficulty - Aptitude Benchmark Scales" - Squadron Leader JOHN W. K. FUGILL, RAAF, Air Force Human Resources Laboratory, Lackland AFB, Texas
1455-1515	Coffee Break
1515-1540	"PTEP-Evaluation Techniques in the Fleet Ballistics Missile Program" - Lieutenant WILLIAM ELLIS, USN, Strategic Systems Project Office, Washington, D.C.
1540-1645	"A Validity Assessment of the Naval Advancement Examinations Through Multiple Discriminant Functions" - Mr. CASIMER S. WINIEWICZ, U.S. Naval Examining Center, Great Lakes, Illinois

THURSDAY, 23 September 1971

0900-0920	"The Pragmatic Approach to Item Analysis" - Mr. HERMAN A. MAHNEN, U.S. Army Enlisted Evaluation Center, Ft. Benjamin Harrison, Indiana
0920-0945	"MOS Mastery Test Development Procedure" - Mr. J. E. HOHREITER, U.S. Army Enlisted Evaluation Center, Ft. Benjamin Harrison, Ind.

THURSDAY, 23 September 1971 - continued

0945-1005	"The Inadequacy of 'Relative' Rating Methods When Size of Ratee Groups is Small" - First Lieutenant KIRT DUFFY, USAF, Air Force Human Resources Laboratory, Lackland AFB, Tex.
1005-1030	Coffee Break
1030-1055	"Mini-surveys" - Mr. ARTHUR G. HERMANSEN, U.S. Army Enlisted Evaluation Center, Ft. Benjamin Harrison, Ind.
1055-1140	"Training Tactical Decision Makers" - Lieutenant Colonel ROBERT E. LOEHE, USMC, Marine Corps Development and Education Command, Quantico, Virginia
1140-1300	Lunch
1300-1335	"Measuring Communicative Skills" - Mr. H. WILLIAM GREENUP, Marine Corps Development and Education Command, Quantico, Virginia
1335-1405	"The Automatic Interaction Detector Among Variables in Personnel Evaluation" - Dr. JANOS B. KOPLYAY, Air Force Human Resources Laboratory, Lackland AFB, Texas
1405-1430	Coffee Break
1430-1500	"Enlisted Job Satisfaction in the Air Force" - Mr. R. BRUCE GOULD, Air Force Human Resources Laboratory, Lackland AFB, Texas; presented by Squadron Leader J. W. K. FUGILL
1500-1545	"The SET Study - a Research Study of the Self-Evaluation Technique" - Dr. JOHN J. HOLDEN, U.S. Army Ordnance Center and School, Aberdeen Proving Ground, Maryland
1830-1930	Reception in the California Room
1930	Banquet in the South American Room Banquet Address: "The Need for New Degree-Awarding Methods" - Mr. JACK N. ARBOLINO, Executive Director, Council on College-Level Examinations, College Entrance Examination Board

FRIDAY, 24 September 1971

0900-0920	"Preference Consistency Testing: Flexible Value Systems and Assessment Reliability" - Squadron Leader BRIAN N. PURRY, RAF, Training Command, Royal Air Force, Brampton, England
0920-0955	"Using Selection Data for Long Term Personnel Evaluation and Planning" - Mr. ARTHUR GARDNER, Senior Psychologist, Ministry of Defence (Navy), London, England
0955-1025	Coffee Break
1025-1115	"An Improved Army Classification Battery" - Dr. MILTON H. MAIER, U.S. Army Behavior & Systems Research Laboratory, Arlington, Va.
1115-1145	"Contemporary Approaches to Validation Research and a Discussion of Seasonal, Regional, and Language Differences in the Canadian Armed Forces Classification Battery" - Mr. HARVEY A. SKINNER, Personnel Applied Research Unit, Canadian Armed Forces, Toronto
1145-1200	Steering Committee Report and Closing Remarks - Colonel L. T. ERICKSON, USMC

TABLE OF CONTENTS

TABLE OF CONTENTS (CONT.)

TABLE OF CONTENTS (CONT.)

OPENING REMARKS

The 13th Annual Military Testing Association Conference was called to order at 0900, 21 September 1971, by Colonel Loren T. Erickson, Head, Personnel Research Branch, G-1 Division, Headquarters Marine Corps.

The various delegates and/or representatives of the several U. S. Armed Forces and of allied countries were introduced, and each made brief remarks. These persons included:

Captain C. E. McMullen	- U. S. Navy
Colonel R. S. Hoggatt	- U. S. Air Force
Dr. R. O. Waldkoetter	- U. S. Army
Commander K. R. Depperman	- U. S. Coast Guard
Squadron Leader J.W.K. Fugill	- Royal Australian Air Force
Lieutenant Colonel R. K. Acheson	- Canadian Armed Forces
Commander Alfredo Beyer R.	- Peruvian Navy

KEYNOTE ADDRESS
BY
LIEUTENANT GENERAL ORMOND R. SIMPSON
DEPUTY CHIEF OF STAFF (MANPOWER)
HEADQUARTERS, U. S. MARINE CORPS
WASHINGTON, D. C.

I am honored to be allowed to address this group today. I want you to know that I stand here as a layman speaking to a group of experts. I'm fully aware that I am speaking now in the only field that I can speak -- as a layman. Maybe some of the things that I tell you from the layman's standpoint may be of some interest in the discussions that will follow. I have read your program and I'm impressed with the breadth and depth with which you are going to explore these things. I think it is most refreshing to have a group of people such as are gathered here: representatives of the U. S. Army, Air Force, Navy, Marine Corps; those interested individuals in the Department of Defense; and our comrades-in-arms from the United Kingdom, Australia, Canada, Peru, Germany; and others -- all joined together in a common cause.

It would seem to me that the overriding purpose of such a convention and such a seminar would be to find a better way to more intelligently use the manpower assets, the only really priceless thing in all of our Armed Services. We all seek a better way to use those priceless assets. It's also refreshing to know that people will join as you're doing here today in a friendly, warm and open atmosphere, in which you can put all the cards on top of the table, in which there are no constraints of classification. In the business in which you are engaged there are no secrets, with the exception, of course, of protecting promotion examinations from compromise. Beyond that there are none of the constraints that sometimes operate in terms of classification, but an open awareness that we can learn from each other, that it's a waste of time to plow a furrow that's already been plowed. There is so little time, that we can't be turning over that same furrow again. We must move on because, believe me, there is little time.

We have got to find a better way to employ--to challenge --the people that are made available to us. Those from the United Kingdom, Canada, Australia and others have already had some ongoing experiences in an all-volunteer

climate, one that we're entering now. I think that the U. S. services in time of war make extravagant use of manpower assets. We really do overall, I think, a rather poor job. We're able to get away with it, if you can say that, or at least to muddle through, because in time of war we have habitually and historically been provided with enormous assets to work with.. When you have a large bank account you can be kind of extravagant and we have been so; and I say this with regret. I don't think the fact that we're able to muddle through should give us any sense of satisfaction. It certainly doesn't me, because we've mishandled a lot of individual people. We have, I guess, enjoyed some measure of success but only because we did have lots of people to work with.

Now, time in that context for the U. S. forces is running out. We are not going to have enormous personnel assets in years to follow. We're going to have far fewer people. I think it's characteristic of all military services that as your manpower resources go down, there seems to be no corresponding diminution in the functions and responsibilities assigned to the individual services. In a very pragmatic fashion, we're going to have to find better ways to employ the reduced manpower assets if we are to even approach the responsibilities, the roles and missions assigned to our various services.

Quite beyond that, we will be forced to do this because we're dealing now, and will be dealing in the years to come, with a different kind of young American than we have faced before. He is one who asks more questions, one who does not blindly follow. However, on the bright side, he is one who will follow effective leadership better than any of his predecessors. But it must be leadership that he believes in and he, himself, must be challenged. The task that we have is twofold. It involves, in the broad sense, finding the better way to employ these people and, as a corollary, insuring that where we place this individual is in a job that he finds interesting and challenging and from which he can get job satisfaction.

We must avoid the situation of blindly poking people into holes to make our charts on the wall in Washington look good, and thereby creating a situation in which a man who might have been a superlative success is placed in a circumstance where he's bored and indifferent and thus becomes an underachiever. This, I think, is the challenge.

Now if I were asked, as I have been asked, to state a keynote for this conference, it would be simply this, ladies and gentlemen, the matter of practicality. I come to you as a layman, asking you, the experts in your fields and in your disciplines, to give us the tools to work with. This is what we need. We need the kinds of tools that will help us solve our ever-growing manpower problems. We need tools to better gauge, to better determine civilian-acquired skills. We need to know more about the relationship of civilian-acquired skills to military requirements, how to bridge those gaps. We need to know how to take a man with a certain civilian-acquired skill, or an aptitude, and to place him in the military environment so as to benefit the organization, and at the same time provide him with an interesting and stimulating challenge, setting out ahead of him goals that he will want to strive for not that he's ordered to strive for but that he'll want to strive for and believe in.

My request to you is for those kinds of tools. You all know, much to your great frustration, that we ask you for formulas, for devices, for gimmicks that will permit us to solve this problem and tie it up in a nice, neat package. I would only ask that you be patient with us in this because you know better than I, that the human equation with which we deal is made up of all variables. There are no constants in that human equation and thus it doesn't integrate very well. The answers that you give us, you often try to qualify and we become impatient with this qualification. I'd like you to know, really, that while we do this we, the laymen, we, the personnel managers, recognize that we do place upon you impossible demands. We also understand that, in the behavioral sciences, in which you are so preeminently qualified, there are no finite answers. You can give us trends, you can guide us in the right direction. You cannot give us nice, neat packages even though we oftentimes demand precisely that.

I also want you to know that we understand thoroughly that in all of these discussions which you'll be engaged in here in the next couple of days and in all of your work that you do throughout the year we are not talking about cold numbers, we're not talking about bits of microfilm or bytes in a computer data bank. We're talking about people--live, real people--individual people with individual hopes and aspirations and fears and concerns. We know that what will challenge one man will not necessarily challenge the next. We really know that there are not fixed answers to our problems, yet we

seek them. If you note, in our discussions and in our requests to you, an edge of desperation, this is because we are so concerned with attempting to solve this problem that we are turning to you, the experts in the field, asking you for help.

Ladies and gentlemen, we really need your help. We need tools to address this problem. We need to look at this idea of motivation, what really is it? How can we achieve this balance that I addressed, of placing the individual, recognizing his civilian acquired skills, bridging the gap from his civilian environment to a military environment, placing him in the area where he can serve best. Then testing, if you can, and giving us the tools that will permit us to move this man through the military environment to higher ranks and continue to provide him with the kind of challenge that will lead him to believe that the work he's doing is useful and needed--work that brings him a sense of satisfaction.

Two years from today in the United States Armed Forces, there will be no pressure of the draft. If we are to have, and we must have, a viable Armed Forces of the United States two years and three years from now, and all the years to follow, these personnel management problems --the idea of placement, the idea of testing for advancement--all of these must have reasonable answers. Today, as your keynote speaker, I ask you to attempt to provide us those tools. You know our requirements, you know that time is short. Our request is for the tools to help us solve these problems. When you bring us tools, tell us how to use them and how not to use them. Tell us of the shortcomings and of the pitfalls in their use. Wise as you are, you cannot devise the perfect human formula. Even through we will continue to seek that, we all know that it's not going to be.

Ladies and gentlemen, we do need help. We must, in the United States Armed Forces, do a far better job of handling our people. We've got to know more about these people and you, who are so highly qualified in the fields of behavioral sciences, are our hope for these kinds of answers. You know that we're looking for the universal answers. You know that we're looking for the universal answers. You know that we're looking for the universal formula and that you can't provide it, but you can come fairly close. You can come closer than any other groups. We ask that you come forward and tell us what can be done. Give us answers for the short term. For the long term indicate the areas in which you think that further

research, further investigation will be beneficial. In doing that, be candid about what may be the outcome. If you can point out to us, in layman's language, what possibly can be achieved by such research and further investigation, then you can depend on our support to convince others that this work should go forward.

What I really ask you for are the practical tools of personnel management, realizing always that we're dealing with live bodies, with individuals and individual hopes. How do we deal with these? How do we challenge them? How do we place them where they can contribute the most? How do we do this in a fashion that convinces them that their work is worthwhile? These are the problems that I see and I solicit your support. I thank you for hearing me out, and I wish you well in your conversations to follow.

SOME IMPLICATIONS OF THE SUPREME COURT DECISION, MARCH 1971 (CIVIL RIGHTS ACT, TITLE VII, 1964)

Raymond O. Waldkoetter
US Army Enlisted Evaluation Center
Fort Benjamin Harrison, Indiana

INTRODUCTION

Upon digesting the opinion delivered by the Supreme Court, the intent of Section 703(h) is made quite pointed to encourage the development and proper use of job-related tests when such are to be applied in personnel selection and classification. This opinion can be used to direct the application and development of occupational tests for business, government and military purposes. There are several implications that occur which will be examined for an initial position of consensus as to any technically defensible group actions that might be suggested during this MTA conference. While reactions to the court decision can be taken individually, there may be merit in devising an accommodation to the opinion which supports the single member organization or program, yet relates it in a functional sense to the actions of the others, depending upon the particular objectives of each program.

The possible implications of the March decision are that some organizations will try to contrive methods to support whatever has been and is being done with their measuring devices, others will make good explanations for testing and ignore their own data, and some will construct better job tests but never really integrate them into the total personnel system. And, each organization can follow one or all of the methods depending upon the technical quality of what has been done and what is being planned.

We should ask what are the alternative approaches to test management, so the purposes of the given program can be fulfilled and the specific legal intent be satisfied according to compatible technical and ethical standards? It is the wisdom of previous trial and error programs such as the power generating company which triggered the court decision showing that testing was only, in that organization, a superficial function providing little benefit in terms of personnel assignment. Advantages and disadvantages do become explicit as we try to define alternatives in response to the legal guidance stipulating job-related tests must be shown to have an actual relationship to performance in a particular job or class of jobs. What course of action regarding appropriate test development can be outlined, to subsume both differing alternatives and integrated actions, to address programs designed for particular objectives but not exclusive from those of other MTA members and kindred organizations?

Since the case of Griggs v. Duke Power Co. (1971) has been used to clarify the test meaning of Title VII of the Civil Rights Act of 1964, there can be no doubt that tests measuring purported abilities must be shown to have a demonstrable relationship with qualifications to identify reasonably proficient performance on the job(s) for which they are used. In this case measures of general intelligence and mechanical aptitude were being used, with requisite scores set for initial hiring and transfers to coincide with the national median for highschool graduates. The only apparent rationale for the testing program was that the power company expected a general improvement in the overall quality of the work force. Naturally, this testing reason did not suffice in view of evidence that those workers who did not complete highschool or take the tests were performing satisfactorily and progressing in departments for which the test

criteria were then being applied. Moreover, since this program had the end result of racial discrimination, the problem of minority group exclusion in job competition can be regarded as the cause in motivating the response of the Supreme Court. Basically, the Equal Employment Opportunity Commission (EEOC) has the enforcement responsibility for interpreting Section 703(h) of the Act, and the guidelines provided permit only job-related tests. The guidelines simply demand that employers have data demonstrating a test is predictive of or significantly correlated with key work behavior requirements relevant to a job or jobs.

In looking at the legislative history relevant in this case, the EEOC's construction of Section 703(h) leads to the conclusion, that employment tests which are judged on the job-related function, does comply with congressional intent. What the congressional meaning is: then any test must measure an individual in relation to the job qualifications and not the individual in the abstract, no matter what the motivation may be, in order that such testing does not disqualify individuals, when job skills are not shown to be particularly deficient.

Although the threat first posed to testing by the case of Myart v. Motorola Co. (1964), now has been largely forgotten, it opened the issue of using any test which might favor one group over another in spite of proven business needs. The proposition of how job related any test may be will continue to give personnel testing many moments of justifiable anxiety in illustrating just how precisely personnel are categorized for assignment and promotion.

THE EEC AND THE ARMY COMMENT

When the Enlisted Evaluation Center was asked to comment (1971) on the impact the Supreme Court opinion might have on the MOS testing program in March of this year, a brief but thoughtful review followed to be certain the Center's measurement policies were in full alignment with Section 703(h) of the Act. We firmly believe and can demonstrate that the great bulk of MOS evaluation tests are related to recognized job requirements which must be met by personnel fitting prescribed job qualifications. The test development procedures are such that the job-oriented test items are prepared by personnel psychologists to be in accordance with the job description, contributions by qualified item writers in a technical setting, a thoroughly prescribed test plan or outline, and referenced job analysis, instructional and other designated materials. Tests are reviewed for content validity by several levels of responsible professionals and technically qualified personnel. Quality control research is conducted to make available data and descriptive information by which predictive or correlational judgments can be made to guarantee relevance to key requirements of work behavior. Detailed item analysis data are produced as well as supporting studies of task content and job performance relationships.

The scope of the Army program, for assessing career personnel in over 900 MOSs on each separate 125 item test for an MOS Code, demands regular and annual reviews to update materials and adjust to Army doctrine and changes in requirements. Obviously, there must be some exceptions where individuals could feel disadvantaged due to the multitude of conditions and assignments. Yet, there have been repeated indications that the program has not only given a greater utility to the personnel inventory of enlisted

skills but has furnished motivation and incentive for job competition and recognition, as seen in the DOD Study of Proficiency Pay (Ogloblin, 1970). There is a terrific number of skills which must be evaluated to assure responsible qualification for an adequate mix of forces, so that properly qualified personnel go to jobs where a favorable range of skills is functionally capable of meeting mission requirements. Thus, the Army can be viewed as an organization dependent upon proven skills where coworkers can rely upon each other's proficiency, otherwise low moral and personnel of insufficient skills predict anything but successful accomplishment of defense objectives.

An ad hoc committee statement from the US Army Behavior and Systems Research Laboratory (BESRL, 1971) has been prepared in draft as a comprehensive Army position on this testing issue. The BESRL has analyzed the Supreme Court decision with no exclusive interpretation as to whether the evaluation procedure "involves a cognitive test, self-description blank, attitude-interest scale, or predictor rating, so long as there can be unbiased evidence for significant validity, either independently or in contribution to a composite predictor. This committee statement ends with the following summary remark:

> Although available research appears to meet the needs to justify present testing for selection and promotion of enlisted personnel in the US Army, it would seem desirable in order to meet the spirit of the Supreme Court decision to provide specific research evidence that minority groups in this population do in fact generally fit the same regression statistics as are found for the overall samples in our prior research. Since such research requires sizeable samples of each minority before trends can be observed, this research would possibly be feasible only with regard to negroes and women and with respect to only selected MOS in each of these groups. In addition, it would be desirable to run check studies of these validities every five to 10 years to allow for changes in characteristics of the population of young men and women and in their relation to training technology. Such studies should also be run whenever there is a significant change in the technical content or training methods for Army jobs or MOSs.

After reviewing the Army's reaction to the DOD inquiry as somewhat typical in regard to the comments made to account for our programs, the question of whether equal employment opportunity provisions of the 1964 Act may even apply to such governmental operations, as come under the DOD, is raised as a natural search for alternatives.

SOME TESTING IMPLICATIONS

Here, then, is the point where the implications of the 1964 Act must be addressed because Executive Orders on Equal Employment Opportunity direct concern on the part of the agencies in the Executive Branch. The implications of how action may be taken to respond to guidelines used by the EEOC brings the issue to examination of the various alternatives that can be operationally described.

One course of action leading from the implication mentioned initially, of supporting whatever has been and is being done with given measuring devices, is that of proving the testing program is meeting its intended

objectives. Within reasonable limits most DOD-MTA members will be able to support their actions by showing the long history that has gone into the development and evolution of current techniques and measuring programs. Where the common objectives in training and job assignment have been to identify degrees of trainability and then assign personnel whenever feasible in the most suitable position, little criticism can be knowingly leveled at the intended results. But the inflexibility that often evolves with well established measurement programs carries an intrinsic constraint, which motivates proponents to defend those program features which are under attack, when these features are well beyond the need for some changes. Assuming that with this alternative the best use of available data is achieved, then the conservative constraint is aggravated yet more by the probable lack of newer categories of data for new problems or personnel decisions, which escape data-oriented policy management.

A second alternative derives from the likely implied reaction where a good explanation of current and additional testing programs may be presented, and all aspects of data gathering will receive adequate coverage with only one minor oversight being encountered. This exception being the partial or serious neglect of important applications of relevant personnel-measurement data. Accordingly, as the personnel-policy maker seizes on the justification of a good system which can be wholly pertinent to personnel requirements, but misses the emphasis necessary for particular data uses, the overall direction of such a program can lead to salient omissions in critical situations. These omissions then will cause a tendency to avoid the applications of data to the solution of touchy

problems, thereby inhibiting the problem-solving function of an organization and discouraging the publication of a data source because its existence could be a trifle embarrassing. With the course-of-action signifying some organizational constraint, there may be a temptation to reorganize selected functional elements, when the answer may have been to simply communicate the required data more directly to the primary user or decision maker in a qualified frame of reference.

A third comparative course-of-action results from the implication felt in constructing better job tests and not having them properly integrated in the personnel system. We may observe that the emphasis could be on job-related measuring devices with the effects never duly assimilated in personnel actions, where the data and information can have a significant impact. In a program that might be struggling to interface measures of aptitudes for trainability and actual job-oriented tests with all sorts of separate and composite data combinations, there will be a greater chance that various data uses will not always be coordinated sufficiently. A firm basis would support any desire to extend the effects such a job-related system could yield. However, since the successful interaction of measuring programs within an overall personnel system cannot always be mutually supporting, there will be some occasions when the job-related tests do not provide scores in a useable context even though the data are truly relevant. What may be inferred is that the attention to job-relevant scores can succeed up to the degree such scores can be accurately interpreted at the various levels of management and operations which have the functional responsibility.

OBTAINING JOB-RELATED TESTS

The briefly discussed alternative actions have been rather broadly outlined, but these may set the stage for a conceivable strategy in guiding an analytical approach for the testing problem. Either a consensus in the administrative and technical treatment of the testing issue or a definitive position can direct a common attitude and methodology in handling the different program needs which evolve in job-related testing. No matter which direction is taken to formulate the best articulation of the administrative and technical treatment of the testing issue with a well defined position, the accommodation to be created among the three alternatives stated above can be discerned by placing maximum emphasis on their overlapping qualities.

To submit a descriptive comment which will contribute more than merely the initial step in constructing a common understanding of the testing issue at hand is to say the least presumptuous. The skills are in the grasp of this combined audience to prepare a clarification of those means which can help produce the developmental actions essential in the effective creation and use of more job-related tests.

At the risk of making the process of doing this seem all too simple, the basic guidance to accomplish the objective is to understand the personnel structure and system of any organization. How its data requirements are met and how it can answer the data and management information questions of the other organizational components and systems. From this standpoint the action should be expedited if the organization can understand precisely how effective its given testing devices are and the data uses being exercised.

Next with a good thorough analysis of the advantages and disadvantages of a particular testing system specified, the decision points where available data are not being applied should be reoriented. A final phase would be opened then with a plan to design those job-related tests that did not occur in prior actions, and to relate career fields and jobs with those validated job duties, tasks, and elements which previously were just assumed for the normally able and proficient incumbents.

Even though the DOD-MTA and other associated members are in a favorable posture to react to any sort of consequences to come from such guidelines as prompted by the EEOC's position, the prevailing trends in occupational and career development research forecast continuing stress and hopefully optimistic achievement in evolving more job-related tests.

REFERENCES

Griggs v. Duke Power Co. Supreme Court of the United States No. 124 - October Term 1970. Washington, D. C.: March, 1971.

Myart v. Motorola Co. Illinois Fair Employment Commission. Congressional Record 5662, 1964 (Reprint).

Obloblin, P. K. Study of Proficiency Pay (Superior Performance). Department of Defense (Manpower & Reserve Affairs): Military Personnel Policy, June, 1970.

US Army Behavior & Systems Research Laboratory. Draft Report of BESRL Ad Hoc Committee on Supreme Court Decision Concerning Testing for Selection and Employment. Arlington, VA: US Army Manpower Resources Research and Development Center, 1971

US Army Enlisted Evaluation Center. Comment on the Civil Rights Act of 1964, Title VII, Section 703(h) and the Supreme Court Writ. Fort Benjamin Harrison, Indiana, 8 March 1971.

USAF Evaluation Systems

Review and Research Recommendations

A Paper Presented to the

Military Testing Association, Sept. 20-24, 1971

by

Robert E. Wilkinson

Personnel Systems Branch
Personnel Division
Air Force Human Resources Laboratory
Lackland AFB, Texas

This paper summarizes the current status of revision of a proposed officer effectiveness reporting system for the Air Force. The final system, based upon specifications established by Hq USAF, is a synthesis of certain findings of panels of experts which will be described.

A workshop of experts drawn from industry, the academic community, government laboratories, and operating agencies of the military forces, the Army, Navy, Marine Corps, and Coast Guard met in January 1971 and created for the Air Force alternative evaluating systems. Participants were divided into five panels and assigned the major task of providing at least one evaluation system proposal with recommendations for follow-on research. In this brief, I will cover the results of the workshop and suggest areas for research.

The participants reviewed evaluation systems that are in use in major organizations: the military, government, industry, and military establishments of foreign countries. You notice that we had coöperation from International Business Machine Co., General Motors, J. C. Penny, the Royal Australian Air Force, the Royal Air Force and all military forces of the United States. Our experts determined that an optimum evaluation system ought to consist of three interrelated sub-systems. These three subsystems are a part of today's Air Force but are limited in precision and scope. The base sub-system is a methodology for precisely describing Air Force jobs and determining job requirements. The second system is focused on determining or defining the performance of the individual in the job and the last system, based on the first two, is a technique for determining the promotion potential of the officer, or how well he is expected to perform in higher grades. Starting at the base of this pyramid and working up, these are the results. The workshop participants were in agreement that ratees should provide an input into preparation of the job description. It was suggested that we overcome the current limitations of the job description system in use; the limitations are great differences in the quality of the job description records, and no standardized procedure for describing the job. Standardizing the guidance that is available to Air Force raters will help eliminate both deficiencies. And having done that they suggest that we apply to the job descriptions, at the rater level, factors which define requirements for each job. These factors are the most important determinants of job requirements based on research conducted at HRL. The factors are: Education, special training, working conditions, the originality, ingenuity and creativity required by the job, communication skills, interpersonal skills, judgment and decision making, planning, management, and risk. A methodology was developed for assigning a numeric value indicating the degree to which the factor is required for the job. Essentially what has been established is a system for describing Air

Force jobs and for translating those descriptions into numeric values that can be used to guide promotion and assignment activities. Our experts next addressed the problem of measuring how well the officer performs in the job. There is a need for an instrument that objectively tells the officer how well he is doing in order that he can improve. Such an appraisal is necessary for adequate career counseling and development. This need is accentuated because officers are assigned across career areas into career areas in which they have had little or no experience. The panels of experts suggested that this instrument be separated from the promotion potential appraisal instrument because keeping them together puts pressure on the rater to inflate both; moreover, the multiple purposes required of an appraisal instrument dictate against a single form. They further suggest that the methodology for job performance evaluation be career area oriented, that attributes evaluated be tailored to different career areas of the Air Force. Suggestions for the job performance format ranged from retaining the current Air Force procedure to modification of Smith and Kendall's scaled expectations methodology. Examples include: (1) Requiring the rater to rate only those factors relevant to the subordinates' present job, the degree of relevancy being indicated by using a five-point adjectival scale. Space is provided for writing in significant unlisted factors. The ratee is then rated on a five-point adjectival scale. No overall score is rated or derived. No word picture is required. (2) For each of the factors rated, the rater would be required to provide a specific statement giving behavioral evidence of the factor rating made. This documentation would be necessary for all ratings not in the middle of a five-point scale. Again there would be no word picture. There was general agreement that performance ratings must be discussed with the ratee. It was further suggested that only assignment personnel (not promotion activities) receive performance forms.

These two sub-systems, job description and job performance established the base for promotion potential evaluation; that is where we are concerned about inflation today. I have distilled the recommendations into two primary alternatives. These alternatives recognize the basic difference in systems against which ratings may be given. The first system is essentially the sort that we use today, that is, have the rater evaluate his subordinate against other Air Force officers of like grade. To get discrimination in this sort of system you must have pressure and control devices that bring about discrimination. Pressure and control devices are available for implementation today. They can be put together into a workable system that should give discrimination. Our experts also recommended that we explore relative rating systems. The rater is required to

rank order his subordinates. This system does not require pressure and control devices as it mechanically gives you discrimination.

I will first address the system that uses pressure and control devices. These are some of the devices that have been identified. Require raters to indicate on each officer effectiveness report rendered the number of ratings completed in a given time period in the past and the distribution of overall evaluation levels rated. Many of the large corperations are using the management committee device to bring pressure on the rater. They require that each evaluation rendered be reviewed by a committee of senior managers. This puts pressure on the rater as he is subject to defending any or all evaluations rendered. Inspector General Offices could also review performance appraisal. Training films are used to inform raters of their responsibility to insure that discrimination is provided. Quota controls, telling levels of command how many ratings of each level can be assigned was suggested. The Australian Air Force and the Royal Air Force are successfully using confidential ratings. Given these pressure and control devices the operating system (of the Air Force) could be refined to incorporate as many of these as are required to get the discrimination that is needed. Some of the features of such a redesigned system could be these: First, the system would add to the rating form a block which requires rater history. It would give feedback to the rater on the distributions of ratings in the recent past so that he would know the frame of reference against which his people will be competing for promotion, assignments and other purposes. The technology is available today for giving feedback to the ratee on his chances of promotion. Under the existing system the ratee does see the evaluation that is given to him but he is unable to determine the probability of his being promoted or his standing among contemporaries. Providing ratees feedback on chances for promotion may remove some of the resistance to making ratings confidential. The system of the future will have to hold the rater and indorser responsible for their actions. Moral persuasion, education, monthly reviews of appraisals rendered or, even IG inspections to determine how raters have rated in the recent past can be used. And finally, if all these are not satisfactory, the system could be confidential although there is no assurance that confidentiality would in itself cause discrimination. The majority of participants recommended varying degrees of confidentiality. In an attempt to introduce variance into the system mechanically, and thereby avoid the moral persuasion or arbitrary management controls mentioned above, elements of the workshop attempted to devise relative rating scales, which would place an officer among his fellows in a comparative way. These systems were found unworkable for various reasons, but I will describe them in the hope that they may stimulate investigations which might solve the reasons for their rejection.

Two relative rating systems were explored by our workshop. The first of these is called the point allocation technique. Its characteristics are described here. Each rater would be given 100 points for each of their subordinates of the same rank. The total available points would be distributed to each of the factors on which the rater evaluates his subordinates. The rater would give between 101 and 130 to above average officers but in order to assign these points he would have to rate other subordinates correspondingly less than 100. This is an open system, the evaluations that are assigned would be discussed with the subordinates. The minimum acceptable pool established was three. If there are less than three subordinates in a pool, then that pool would be merged with pools in other organizations under a common supervisor; and, that supervisor, the second echelon supervisor, would be required to assign ratings based on personal knowledge and supportive information by the first level supervisor. Feedback to the ratee and the rater is provided by reporting the ratees probability for promotion or standing with contemporaries and the overall distribution of scores for the raters. In addition to this kind of relative rating system the workshop identified a free scale no tie system commonly called FRESCA which uses a scale ranging from 0 to 99. The rater would be required to assign to each of his subordinates a value along the scale for each factor evaluated. The system would follow the rule of no ties which in effect means the supervisor would be required to rank order his people. The maximum ratee pool would be ten because the rule of no ties would require that the maximum value which could be assigned to the low man would be 100 minus the number of people in the ratee pool. Hence, large ratee pools would tend to discriminate against the low individual. Ratings would be assigned on the overall evaluation of the promotion potential of the officer as well as on a number of sub-factors. The workshop recommended that this system be confidential. We expect that there would be inflation in the centile scale without confidentiality although control and pressure devices outlined previously could be incorporated. Feedback would be provided to the ratee at critical decision points on his probability of being promoted or standing among contemporaries. Raters would also be advised of the current distribution of scores. Under the FRESCA system the rank ordering is available if inflation should develop in the centile scale. The panels of experts were aware that pools in the Air Force are not big enough and that rules for forming larger pools must be determined. We could have the supervisor rate or rank order not only his own subordinates but others as well. This might tend to dampen the enthusiasm by which subordinates perform jobs for immediate supervisors and it might undermine the authority of that supervisor. Another alternative that can be evaluated is that of having subordinates evaluated in a population of people whom the rater has rated in the past. The greatest danger is that raters forget who they rated and

are not able to discriminate well between past and current subordinates. Another alternative is to have the ratee rank ordered with all the supervisor's subordinates regardless of grade. This presumes that they are all competing for the same grade and that is not correct. They lack common experience and common development. Finally, we can explore the possibility of having small ratee pools merged with other ratee pools and the individuals in them evaluated by common supervisor at a higher level of supervision. With that organizational distance between the rater and ratees there is logical concern as to whether meaningful discrimination could be achieved. In addition to the problems in forming ratee pools, FRESCA and the PAT system share two common problems. One is differences between the quality of officers of different ratee pools. The assignment system intentionally assigns quality officers to particular organizations which suggest a need for methodology to discriminate between pools. The second problem is inequalities created by different size pools, necessarily, there is a greater probability of error in individual rating in small pools.

The systems described pull together many of the workshop recommendations, however, there were a number of topics that have evolved from the workshop, and from our experience in attempting to build the evaluation system, which are in need of careful investigation. They are as follows:

1. Develop features to include in the job performance form for various occupational areas.

2. Develop methods to expand ratee groups.

3. Methodology for computation of differential promotion composites (by occupational field).

4. Develop assignment composites through differential weighting of performance and potential factors.

5. Determine frame(s) of reference for raters and ratees in the rating process.

6. Determine whether the method of submission has effects on completed OER.

7. Determine whether feedback on rating trends provide measurable improvement in rater action.

8. Determine whether independent rating judgments on promotion potential by raters and indorsers provide more accurate rating estimates.

9. Develop methods and controls for training raters on objectives and techniques of evaluation.

10. Identify feedback information that secures acceptance of users regarding the reporting of promotion and career decision.

11. Develop career advisory procedures and/or rewards needed to maintain performance standards of personnel not selected for promotion.

12. Do pressure(s) of the need for diverse specialization of skills demand an evaluation system that will keep compensation compatible with general economic conditions?

13. Explore selection board actions to determine the most equitable procedure.

14. Study desirability and method for best informing ratees of their probability of success, including predictions of probability of promotion, relative standing indices, and personal review of performance jackets with or without counsel of career monitors.

15. Develop an individual utility index related to probability of promotion incorporating supply, demand and income level indices.

16. Analyze various types of inflation controls; e. g., statistical, managerial, training, and organizational.

ANALYSIS OF ENLISTED EFFICIENCY REPORT TRENDS

Kenneth C. Liebfried
US Army Enlisted Evaluation Center
Fort Benjamin Harrison, Indiana

At the MTA Convention in New York City in 1969, Mr. John A. Burt presented a paper on the redesign of the US Army's Enlisted Rating Form, the Enlisted Evaluation Report or (EER). The form discussed by Mr. Burt was adopted by the Army and implemented into the Army's Personnel Management System on 1 July 1970. The scores obtained on this form became part of the enlisted personnel's MOS evaluation scores beginning with the February 1971 MOS evaluation period.

Slide 1 on

Let us now take a look at the EER and briefly review its design and scoring procedures. Part 1 of the EER contains personal data about the rated individual. These data are entered on the form by the ratee's personnel officer. Part II is completed by the rater. Part IIA asks for a description of duties performed by the individual which are not included in the job description of the person being rated. Part IIB contains six characteristics which are to be rated. This part of the form counts for 80% of the EER score. Part IIC asks the rater to indicate the advancement potential of the ratee. The rater is to assume that he has the authority to promote the ratee and that the ratee will continue to work for him. This part of the form counts for 20% of the EER score. Part IID asks the rater to recommend career development information concerning the ratee while part IIE provides space for the rater to comment on any part of the rating he has given. Part III of the form is completed by the reviewing officer. The reviewing officer may indicate his concurrence or non-concurrence and the reasons he does not concur.

Slide 1 off

When the EER was implemented into the Enlisted Evaluation System, the submission of the forms was also changed. Under the previous system, the EER was submitted at the time the ratee was to take his MOS evaluation test. With the implementation of the present EER, each soldier is rated a minimum of twice a year. Additional ratings can be submitted under special prescribed circumstances. All EERs received are weighted, combined, and averaged. This becomes the individual's Enlisted Efficiency Report Weighted Average or EERWA. The EERWA on file at the time of MOS evaluation testing is utilized in the computation of the individual's evaluation score. By having a current EERWA available for all enlisted

personnel, the Army can use it for other personnel actions such as selection for schools or for duty assignments.

The question this paper will address is: "How has the EER performed as a rating instrument?"

Slide 2 on

Slide 2 shows the comparison of the EER data through July 1971 and the last data available for the previously used EER. In all pay grades the means are lower and the dispersion is greater for the present EER. It can be seen that there is approximately a ten point spread between pay grades. It should be noted that the maximum score for the EER is 125, while the maximum score for the previous EER was 116. While inflation exists in the upper pay grades of the EER, it is not nearly as great as with the old EER. Pay grade E-9 for the old EER had a mean of 115.1. This is less than one point below the maximum score. For this pay grade the differention among examinees would have been made almost entirely on the MOS evaluation test score. The less inflated scores on the EER provide more spread among soldiers and allows the Army to make more meaningful personnel management decisions based on EERWA scores.

Slide 2 off

Slide 3 on

Slide 3 shows how the pay grade means have varied from month to month since the first data were made available. It can be seen that there has been very little variation for any pay grade except for a slight increase in the lower pay grades for the month of May. It is hypothesized that this increase occured because the "Army Times" reported that soldiers not measuring up on their evaluation scores might not be eligible to re-enlist. These means leveled off in June and July and are about the same as they were before May.

Slide 3 off

Slide 4 on

Slide 4 shows pay grade means for Part IIC of the form, "Advancement Potential." The same stability is seen as on the previous chart, however, the means are somewhat lower for this part of the form. Charts for each rating characteristic are being maintained at the USAEEC. Each Characteristic shows the same stability as the slides shown here today.

Slide 4 off

In review, the data indicate that the EER has been quite stable during its first year as a part of the Army's personnel management system and that it is less inflated than the previously used EER. The data received are very encouraging, a few areas that might be modified to further improve the form are presently being investigated.

Scoring results indicate that the Advancement Potential Section of the form receives lower average scores than the characteristics section. In many cases the rater is giving ratings of outstanding and excellent on all of the characteristics, yet is checking the "promote with contempories" box in the Advancement Potential Section. These inconsistencies tend to indicate that when a hard decision about the ratee's future has to be made, the raters may be giving more realistic ratings. These inconsistencies have also caused some inquiry by enlisted personnel who thought they received a "good" rating, but whose reported EERWA was below the score they believed they should have received and in a few cases their scores were below the average for their pay grade. Individuals who have fallen into this category are blaming the form for their lower than expected scores, while the real culprit is rater inconsistency. The rater might believe the rating he gave to be an excellent one because (1) he did not read the back of the form where the advancement potential is explained, (2) he was not able to disassociate old thought patterns about promotion which included time in grade and other requirements or (3) he might honestly believe that to promote with contemporaries is a better than average rating.

A second area that has caused some inquiries were from enlisted personnel in the upper pay grades who received average or above average ratings on both parts of the form and then received an EERWA below the mean for their pay grade. The possibility of this occuring was considered when the form was being designed, however, it cannot be eliminated as long as there is inflation of scores in the higher pay grades.

Another interesting occurrence on some forms is the ratings given and the comments about the ratee do not agree. This does not affect the score received, but it causes one to wonder what the rater was thinking about when completing the rating. These individuals generally receive ratings in the average or above average boxes, but are described as being outstanding performers.

<u>Slide 5 on</u>

Slide 5 is an actual EER that shows the various rating inconsistencies I have mentioned. Five of the six characteristics were rated outstanding and the sixth one was rated excellent. The advancement potential is promote with contemporaries and the comments indicate the ratee is excellent.

<u>Slide 5 off</u>

The questions which have been asked the Center about the scoring of the form and the reporting of the scores have not been of such a nature to cause any real alarm, rather they have served as starting points for a look at possible modifications of the scoring system or the reporting of EERWAs to individual enlisted men.

Two systems for reporting the scores have been investigated. The first would be to convert the EERWA to a score on the Army Standard Scale. This would provide the enlisted man a scoring system that he is used to using and it would provide him with a good idea of how he compares with his contemporaries on the EER. This system would compress the scores somewhat; from 0-125 on the EER to 40-160 on the Army Standard Scale. Implementation would require a significant resource expenditure, both in computer time and manpower to produce data of limited use.

Another approach would be to report EER results as percentile scores. This procedure would provide the enlisted personnel with definitive information as to how he compared to his contemporaries. As in the previous approach, the expenditure of resources compared to the benefits gained are excessive. There would be a greater compression of scores using this procedure, from 0-125 to 0-100. This would not solve the problem of an individual receiving an average EER numerical rating and being below the mean for his pay grade.

The idea of changing the weights of the characteristics and advancement potential sections of the form is presently being analyzed. Various weighting techniques are being applied to the EER to determine the effects on the individual scores and the rank order of the distribution. This analysis is not yet complete, therefore, I can provide no data to you at this time.

None of the procedural changes I have discussed are presently being considered for adoption into the enlisted evaluation system. They are considerations of changes that might be made if it is determined at a later date that the EER is not as an effective rating as we believe it to be at the present.

The best rating form is only as good as the personnel using the form make it. A chart shown earlier indicated an increase in EER means for the lower pay grades the month the "Army Times" reported that personnel not measuring up to certain standards might not be able to reenlist. Any change of this type to the personnel management system can affect the results reported on the EER. If DA were to announce that there was to be a fifty percent reduction in certain MOS codes, I would expect the EERs for individuals in these MOS to be lower than their previous MOS scores. On the other hand if it were announced that certain MOS codes were well below strength, I would expect individuals in these MOS to receive higher EER scores than in the past.

No rating form will compensate for rater leniency, halo effect, or changes to a system which can affect the ratings being given. Whenever rating scores become inflated the tendency is to blame the form. This is quite unrealistic.

Because the trends discussed earlier in this paper indicate that the EER has been stable for a year, we believe it is a very satisfactory rating form. Therefore we have decided to attack the problems that arise in the use of rating forms by trying to educate the rater in the purpose of rating forms, what he should consider when making his ratings, and explain the various types of rating errors that are common to all rating systems. We could even point out that there is a significant difference between ratings given by NCOs and by officers. An enlisted person who believes he needs a higher rating is advised to convince an officer to rate him.

The exact approach and details for training raters has not been decided upon. The possibility of a training film has been discussed. It is hoped that by training the raters, the constant revision of rating forms can be slowed down. Not every rating form that has been used has been a "good" form, however, many of them which were at least adequate have been discarded because the raters did not understand the rating system or system changes caused the form to be used for purposes it was not designed for. We at the USAEEC believe it is time to train personnel in the use of ratings rather than to attempt to do the impossible in trying to design a rating form that will correct for inflation, halo, and systems changes.

In summary, the EER has been very stable during its first year as the Army's enlisted rating form. A few questions have been raised and these have been or are now being studied and analyzed. It has been decided to develop some type of instructions for the raters which we hope will lessen the rating errors present in our rating systems.

ENLISTED EFFICIENCY REPORT

(AR 600-200 and AR 135-205)

PO INITIAL

PART I PERSONAL DATA (To Be Completed By Personnel Officer)

A. NAME, RANK, ORGANIZATION and STATION

B. DUTY POSITION TITLE

C. PMOSC

D. SMOSC

E. DMOSC

F. DATE OF RANK

G. PAY GRADE

1 2 3 4 5 6 7 8 9

H. DOES EM HAVE OVER 3 YEARS' SERVICE?

YES NO

I. TYPE OF REPORT

REG SPEC CDY CR PCS OTHER

J. SSAN

0 1 2 3 4 5 6 7 8

K. PERIOD OF REPORT

BEGINNING MONTH: J F M A M J J A S O N D

BEGINNING YEAR: 70 71 72 73 74 75 76 77 78 79

ENDING MONTH: J F M A M J J A S O N D

ENDING YEAR: 70 71 72 73 74 75 76 77 7

PART II RATER (To Be Completed By Rater)

A. BRIEF DESCRIPTION OF DUTIES OR RESPONSIBILITIES NOT INDICATED BY DUTY MOS. IF ANY.

B. CHARACTERISTICS	O	E	AA	A	BA	U	RV
1. ADAPTABILITY							
2. ATTITUDE							
3. INITIATIVE							
4. LEADERSHIP							
5. RESPONSIBILITY							
6. DUTY PERFORMANCE							

C. ADVANCEMENT POTENTIAL.

IF I HAD THE AUTHORITY AND RESPONSIBILITY TO DO SO I WOULD:

PROMOTE EM IMMEDIATELY		PROMOTE EM AHEAD OF CONTEMPORARIES		PROMOTE EM WITH CONTEMPORARIES		NOT PROMOTE EM	DENY EM CONTINUED ACTIVE DU-
1	2	3	4	5	6	7	8

D. RECOMMENDATIONS FOR CAREER DEVELOPMENT (Not Counted In Score)

	YES	NA
1. CONTINUE IN CURRENT DUTY MOS AT PRESENT ORGANIZATIONAL LEVEL.		
2. ASSIGNMENT IN CURRENT DUTY MOS AT HIGHER ORGANIZATION LEVEL.		
3. ASSIGNMENT IN DIFFERENT DUTY MOS. (IF YES, SPECIFY MOS ____________)		

	YES
4. ADVANCED MOS–ORIENTED SCHOOLING (IF YES, SPECIFY MOS ____________)	
5. DA NCO DEVELOPMENT COURSE.	
6. SELECTION FOR CIVILIAN SCHOOLING.	

E. COMMENTS OF RATER (Brief Specific Comments, Limit To Space Provided)

F. RATER'S ORGANIZATION AND DUTY ASSIGNMENT

G. NAME AND RANK

H. RATER'S SIGNATURE

I. DATE

PART III REVIEWER (To Be Completed By Reviewer)

A. THIS REPORT WAS PREPARED BY CORRECT RATER. I CONCUR WITH RATER

I DO NOT CONCUR WITH RATER FOR THE FOLLOWING REASONS:

B. REVIEWER'S ORGANIZATION AND DUTY ASSIGNMENT

C. NAME AND RANK

D. REVIEWER'S SIGNATURE

E. DATE

INSTRUCTIONS FOR COMPLETION OF THE ENLISTED EFFICIENCY REPORT

GENERAL INSTRUCTIONS. The Personnel Office is responsible for completing PART I. The Rater will complete PART II and the Reviewing Officer will complete PART III. Read all instructions before completing the form.

The Commanding Officer of the individual being rated is responsible for designating the rater. Raters will be in the direct line of supervision over the rated individual, serving in pay grade E-6 or above and at least one pay grade higher than the individual being rated. The reviewing officer will be a warrant or commissioned officer, normally the first in the direct line of supervision. Only those exceptions noted in Paragraph 8-5, AR 600-200 are authorized.

INSTRUCTIONS GOVERNING THE USE OF THE MARK SENSE PORTIONS OF THE FORM (PART I, SECTIONS G, H, I, J, K; PART II, SECTIONS B, C, D; PART III, SECTION A). Use soft pencil only. Be sure each mark is black and completely fills the space inside the box you wish to mark. Make sure the marks do not extend outside the box. Erase completely any mark you wish to change. Do not fold, tear or otherwise mutilate the form.

INSTRUCTIONS FOR PREPARING PART I (Refer to AR 600-200).

Enter initials only after entire form has been completed, signed by both rater and reviewer, and checked for completeness.

Section A. Print or type the rated individual's name, rank, organization, and station (e.g., DOE, JOHN A.; SFC; 5/11, 6th Inf. Div.; Ft. Jones, Nebraska 34026). The Standard Personnel Plate may be used.

Sections B through F. Enter the required information for each section in the spaces provided (Enter exactly as recorded on DA Form 20).

Sections G, H, and I. Mark the appropriate box in each section.

Section J. Enter the rated individual's SSAN in the left-hand column. Enter the numbers vertically, from top to bottom. DO NOT lay form on side and enter from left to right. After numbers are entered, mark the corresponding box to the right of each number.

Section K. Mark the boxes for both the month and year (last two digits only) of the beginning date and the ending date of the period covered by this report.

INSTRUCTIONS FOR PREPARING PART II.

Section A. Complete this section only if those duties performed by the rated EM differ from those normally associated with the duty position title in PART I, Section B.

Section B (Characteristics). Care and attention must be directed toward marking the most accurate and reliable ratings possible. Rate the EM on each of the six characteristics described below. You may explain significant strengths and weaknesses in Section E (Comments). Mark your ratings in SOFT PENCIL on the basis of the following scale and then enter IN INK the selected abbreviation in each Rating Verification (RV) block in the column to the right of the double line.

- O = Outstanding – performs better than any soldier you know
- E = Excellent – performance equaled by very few other soldiers
- AA = Above Average – performs better than most soldiers
- A = Average – performs as well as most soldiers
- BA = Below Average – performance meets only minimum standards
- U = Unsatisfactory – performs in an unsatisfactory manner

1. *ADAPTABILITY.* Rate the EM on his ability to be flexible and adjust to changing work demands. Consider the presence of mind he possesses, his ability to grasp new concepts and ideas, and ability to analyze and solve complex working situations. Evaluate the EM's capacity to maintain a proper perspective in situations by his sound judgment, creativity, and resourcefulness.

2. *ATTITUDE.* Rate the EM on the degree to which he displays the cooperativeness, sincerity, and interest necessary to maintain proper relations with subordinates and superiors. Courtesy, dignity, and willingness of the soldier reflect standards of conduct consistent with the spirit of the chain of command. Assess the morale of the rated EM and his participation in the mission of his organization.

3. *INITIATIVE.* Rate the EM on his energetic application and attention to duty. Judge the efforts of self-improvement, ambition, and motivation displayed by the EM, as well as the drive and force he demonstrates.

4. *LEADERSHIP.* Rate the EM on the positive manner with which he makes decisions and the confidence he places in them. Consider the ability of the EM to influence or direct the actions of others while maintaining their loyalty. Also, consider the ability of the EM to plan, organize, coordinate, and assign work and the aggressiveness with which he carries out his mission.

5. *RESPONSIBILITY.* Rate the EM on his integrity and willingness to accept the responsibility for his own actions and the actions of others in his charge. The authority he assumes and the judgments he must make should result in complete and economical performance of duty. A responsible EM will possess high standards of military behavior and performance.

6. *DUTY PERFORMANCE.* Rate the EM regarding his overall duty performance and skill. The efficient, thorough, and conscientious production of an acceptable quantity and quality of work is important. Consider the complexity, range of knowledge required by the job, and the reliability and dependability of the rated EM.

Section C (Advancement Potential). For purposes of this rating, assume that you have the allocation, responsibility, and authority to promote this individual and he will continue to work under your supervision indefinitely. Rate the EM on his ability to perform in the next higher grade by considering his total capacity, strengths and weaknesses in comparison with other individuals of his grade and length of service. Descriptive statements are provided for five of the boxes. If you believe that the individual falls between two of the descriptive statements, mark the box in the space between the two statements. Mark your rating in SOFT PENCIL and then enter IN INK the number of the rating you selected in the RV box to the right of the rating section. (In determining what rating to assign a soldier in pay grade E-9, consider his potential for advancement to a higher level of responsibility.)

Section D. If the box titled OTHER has been marked in Part I, item I, DO NOT complete this section. This section provides you with an opportunity to recommend future assignments and schooling for the rated individual based upon your estimate of his ability to assume greater responsibility and/or benefit from additional schooling or assignment. Although the recommendations are not scored, they are extremely important to career management. Full consideration must be given to the requirements of different assignments and the final responsibility involved. You should also consider carefully the rated individual's present level of experience, his capability for development, and his suitability for his present duty MOS. When considering recommendations 4, 5 and 6, select the one type of schooling which would be the most beneficial to the individual if provided as the next assignment. Each statement in section D must be marked yes or NA, as appropriate.

Section E. The purpose of this section is to provide for brief narrative comments on the manner in which the rated soldier has performed his present duties or to explain those ratings which you believe need supporting remarks.

Sections F through I. Self-explanatory.

INSTRUCTIONS FOR PREPARING PART III.

Section A. It is the direct responsibility of the reviewing officer to insure that the proper rater has completed the EER and that an accurate and objective rating has been prepared. Guidance for this determination may be obtained from the ratings themselves, the rater's recommendations for career development, and from the rater's comments. If the reviewing officer agrees with and indorses the ratings awarded, the CONCUR box will be marked. If the reviewer nonconcurs, and cannot reconcile differences of opinion with the rater, the NONCONCUR box will be marked and an explanation of the basis for nonconcurrence will be made in the space provided.

Sections B through E. Self-explanatory.

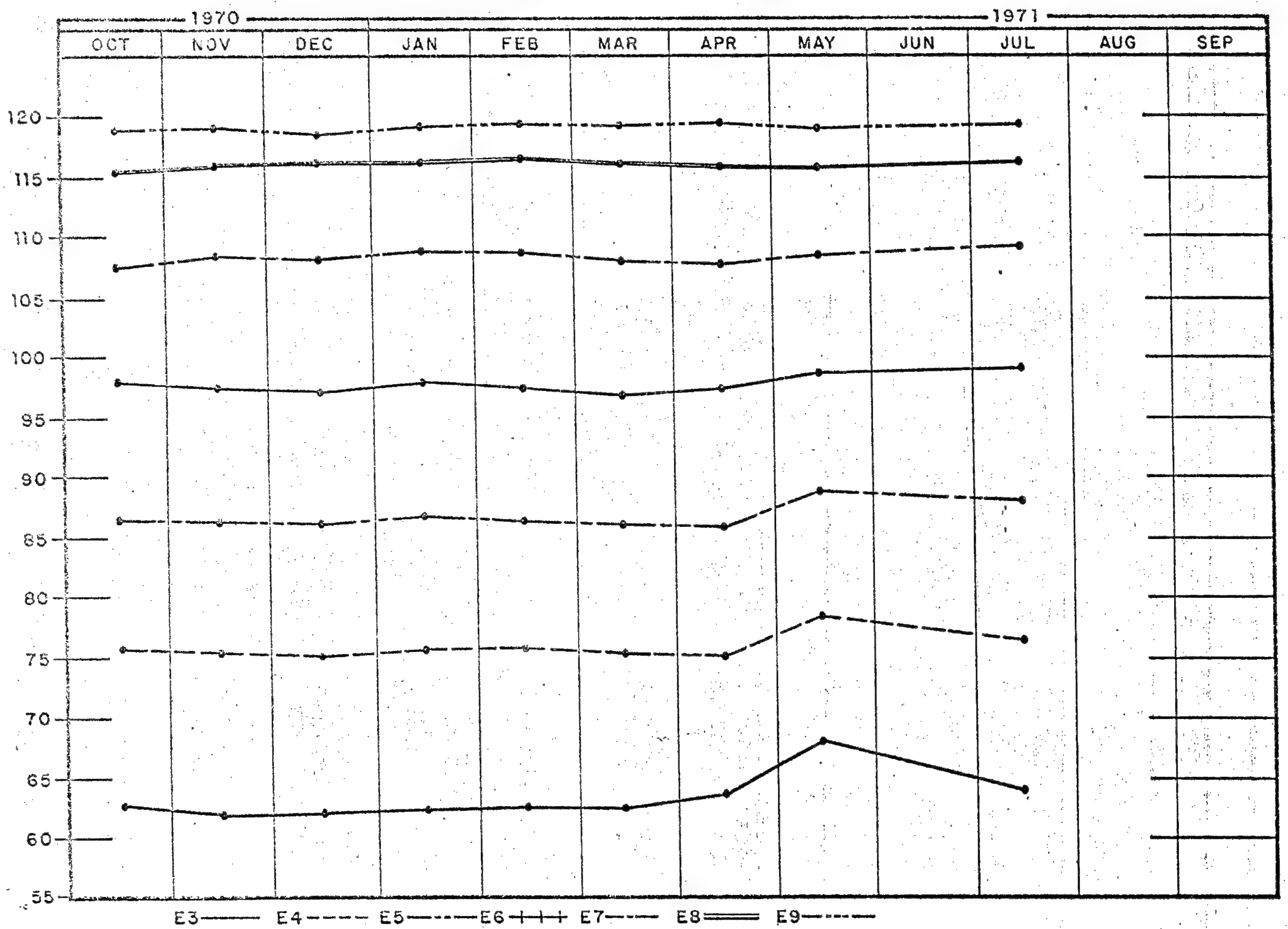
COMPARISON OF EER SCORES BY PAY GRADE
JULY 1971 vs DECEMBER 1970

Pay Grade	July 1971 $\bar{X}$	July 1971 SD	December 1970 $\bar{X}$	December 1970 SD
E-3	63.59	32.12	78.9	28.26
E-4	76.75	27.93	95.9	20.47
E-5	88.17	25.07	102.8	16.09
E-6	98.73	22.05	107.5	13.17
E-7	108.90	17.41	111.8	9.28
E-8	116.06	13.88	113.6	6.92
E-9	119.57	11.45	115.1	3.76
Overall	99.12	23.99	105.1	15.99

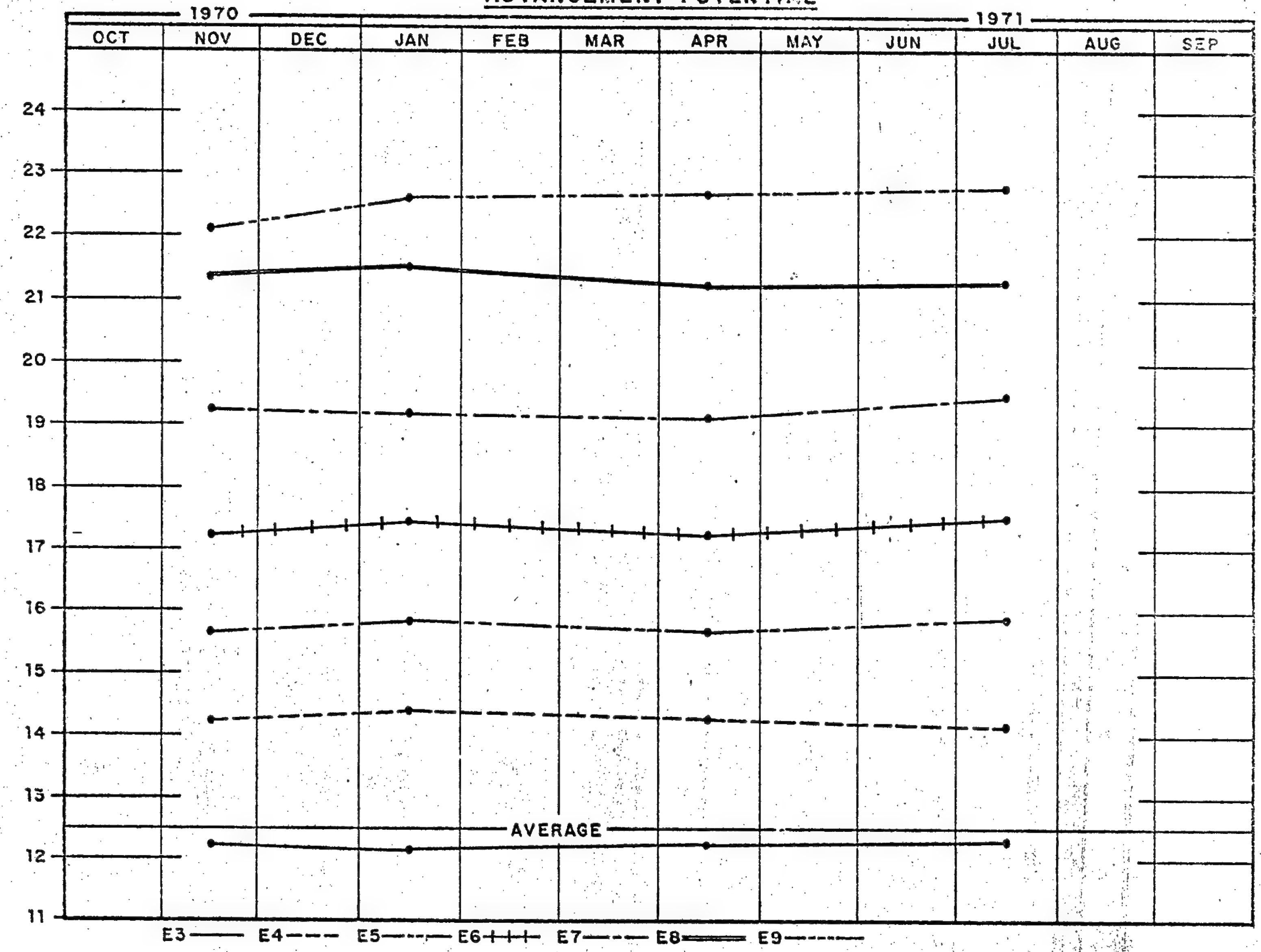
Maximum Score for present EER = 125
Maximum Score for previous EER = 116

Slide 2

EERWA MEAN BY MONTH



ADVANCEMENT POTENTIAL



RACTERISTICS	O	E	AA	A	BA	U	RV
PTABILITY	■						O
TITUDE	■						O
TIATIVE		■					E
ADERSHIP	■						O
SPONSIBILITY	■						O
TY PERFORMANCE	■						O

C. ADVANCEMENT POTENTIAL.

IF I HAD THE AUTHORITY AND RESPONSIBILITY TO DO SO I WOULD:

PROMOTE EM IMMEDIATELY		PROMOTE EM AHEAD OF CONTEMPORARIES		PROMOTE EM WITH CONTEMPORARIES		NOT PROMOTE EM	DENY EM CONTINUED ACTIVE DUTY	RV
1	2	3	4	5	6	7	8	
				■				5

COMMENDATIONS FOR CAREER DEVELOPMENT (Not Counted In Score)

	YES	NA		YES	NA
NTINUE IN CURRENT DUTY MOS AT ESENT ORGANIZATIONAL LEVEL.	■		4. ADVANCED MOS—ORIENTED SCHOOLING (IF YES, SPECIFY MOS ____)		
SIGNMENT IN CURRENT DUTY MOS AT GHER ORGANIZATION LEVEL.			5. DA NCO DEVELOPMENT COURSE.		
SIGNMENT IN DIFFERENT DUTY MOS. YES, SPECIFY MOS ____)			6. SELECTION FOR CIVILIAN SCHOOLING.		

MENTS OF RATER (Brief Specific Comments, Limit To Space Provided)

Sg [redacted] has been NCOIC of the radiology dept. for the past 12 months, as done an excellent job both in administration and technically. He commands e respect of his men.

AUTOMATED TESTING AND ATTRITION
CONTROL (ATAC II)

Presented By

DAVID A. DEORE

Marine Corps Communication-Electronics School

Twentynine Palms, California

INTRODUCTION

The Marine Corps Communication-Electronics School is a formal school organized to provide training of operators and technicians for Marine Corps Ground and Aviation Ground Communication-Electronics Systems. The School is located at Marine Corps Base, Twentynine Palms, California with a Sub Unit at Marine Corps Recruit Depot, San Diego, and has a student load of 5000 students annually. In the recent past, the yearly student totals have been in excess of 10,000 students. The School teaches 50 separate courses, which constitute a total of over 17,500 hours of course material. As many as 4000 students might be tested in a single week. The number of students involved require steps to insure that quality instruction is continually providing critically needed trainees to the field. Since failure to provide adequately trained personnel would be of tremendous consequence; the testing routines employed to ensure quality are of special interest to the Marine Corps Communication-Electronics School. It has been apparent for some time that an integrated system of automated testing would be of particular interest to the School.

HISTORICAL DEVELOPMENT

Experimentation with such an automated testing system began in 1963. It was felt at the time that automation was needed

in achieving a reduction in personnel and time to grade and analyze test scoring and question construction. The maintenance of academic history also appeared a likely area where automation would conserve time and effort. The Automated Testing and Attrition Control System was then conceived under the direction of Dr. Richard S. HATCH. The System was written in IBM 1401 programming language and included provision for student grading process, academic history, and test analysis and development. ATAC I was developed for the Electronic Fundamentals School, which was the only School where the attrition control subsystem was fully implemented. It was the attrition subsystem based on the normal curve that later proved to be unsatisfactory in light of the School's training objectives. Several reports were generated by ATAC I. They included the Attrition Control Summary Report, the Student Answer Card Listing, Grade Reports and History and Analysis Reports. One of the interesting aspects of ATAC I was the ability to create the actual tests by selecting test questions from a question bank in a pattern influenced by the test analysis data. This aspect of ATAC I was retained when the system was updated and may prove to be one of the main advantages of the system. About a year and a half was spent in the development of ATAC I, and much valuable information was gained through experience with the system. In 1969, however, it was clear that the system needed revision. Primarily, the system had been intended for use with the Electronics Fundamentals

School and a wider application was desired. No formal documentation existed on the system and training Marine Corps personnel without it would be difficult. Run time on the 1401 computer was in excess of that desired, and conversion to the IBM 360/30 would improve that area of system operation. Many other technical improvements were also needed in system input/output and provision for graduate evaluation was lacking. On 11 August 1969, a contract was awarded to Computer Application, Inc to revise and update ATAC I and to implement an ATAC II System. The contract was to be 6 months in duration. Due to many factors the contractor, a national concern, went bankrupt in September of 1970. 1 October 1970 Systems Consultants, Inc. bought the contract and became the prime contractor.

II. ATAC II

ATAC II was developed to provide a fast, accurate, and economical means of performing the administrative functions related to academic testing, analysis, and control. This section describes ATAC II, the functions it performs, the inputs required, the processing of the inputs, the outputs produced, and the educational theory used in the selection of the outputs and processing algorithms.

ATAC II performs five major functions: test scoring; student grading and academic record keeping, test and individual test question statistical analysis, item selection and test preparation, and graduate evaluation. Figure II-1 illustrates the overall operational flow of ATAC II. ATAC II is operational on both the 360/30 and the 360/40 computers. The programs are written in COBOL and JCL.

Test Scoring

Students record their weekly test answers on punch cards which are input to ATAC II. The test scoring process consists basically of reading student answer cards, editing each card for errors, scoring each answer card, and computing student scale and standard scores.

Editing of the student answer cards is performed in order to minimize the possibility of entering erroneous data into either student records or historical item records used for statistical analysis. Both the key fields, which identify the test, class and student, and the answer field are edited. Whenever an error is detected in the key fields, the error is listed on an error report and the data on the answer card is not entered into the file. The data from an answer card with an error in the key field is not entered into the file because answer data is sorted by test, class and student for various scoring, grading and record keeping computations. An error is one of these identification fields would cause the answer data to enter some of the computations and be excluded from others

thereby causing inconsistencies in the records and biasing the statistical analysis.

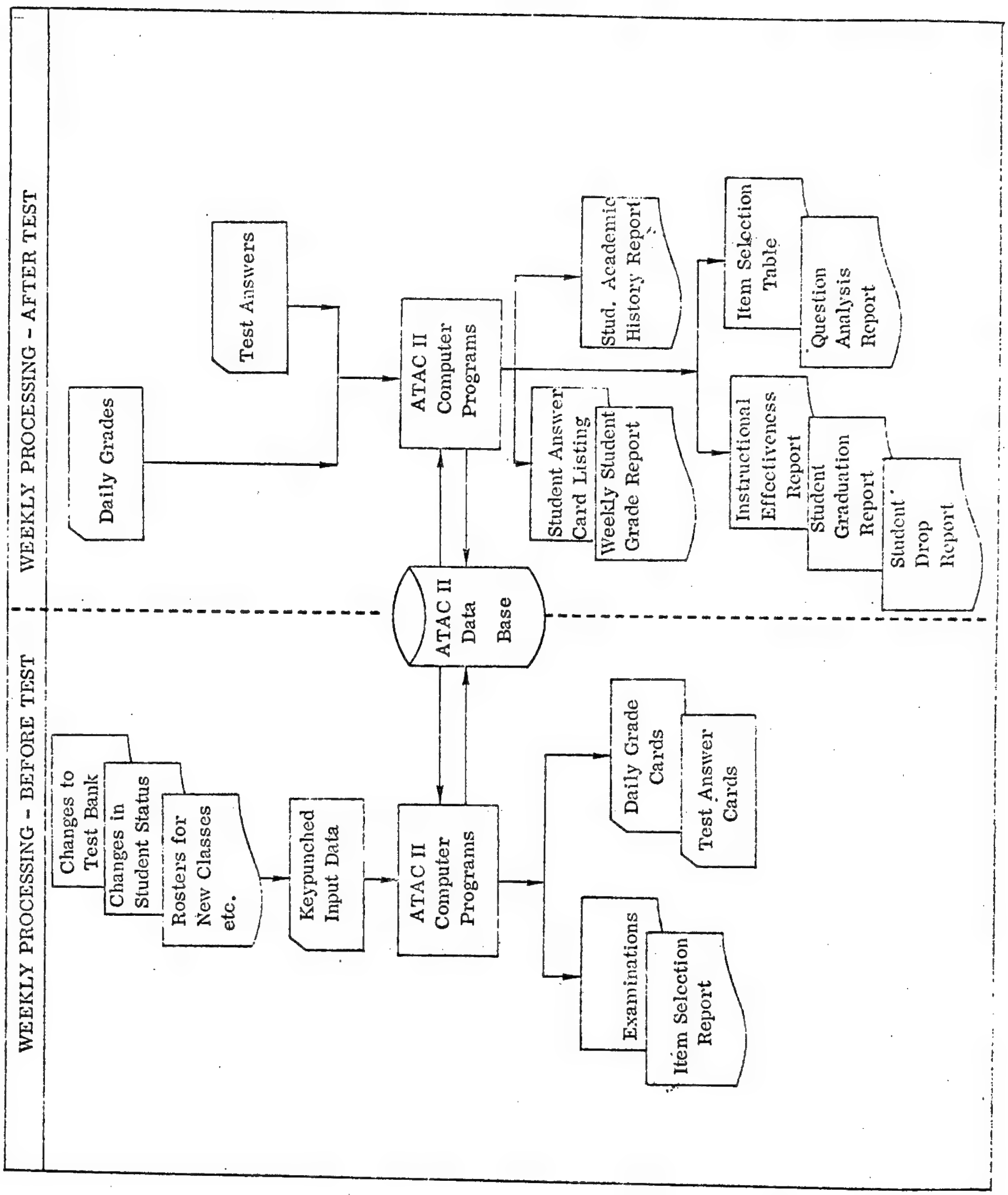
Whenever an error is detected in the answer field, it is recorded as an invalid answer and the answer card data is processed in the normal manner since the scoring, grading and recording keeping computations take into account invalid answers.

Student raw percentile scores and item difficulty levels are corrected for guessing by subtracting "one/number of answer choices" from the student score and the item difficulty level each time a student gives the incorrect answer to an item. This assumes that a student who doesn't know the answer to a question will select one of the possible answers at random. Therefore the student has one chance in the number of choices to guess the correct answer (e.g., on a question with a choice of 5 answers there is a 1/5 or 20% probability of guessing the correct answer).

Student ability follows a normal distribution; therefore, student raw percentile scores tend to cluster close to the mean or average score. In order to spread student scores uniformally on the 0 to 100 scale, raw percentile scores are converted to scale scores using the normal transformation truncated on the 0 to 100 scale.

Each test has a different difficulty level due to differing questions making up each test. In order to eliminate the bias of varying test difficulty levels and to compare a student's ability to all previous students instead of to only students taking the same test, scale scores are converted to standard scores. This is done by first computing the expected value and standard deviation of scale score for the test using past history on each item in the test. Then the difference between the student scale score and the test expected scale score is divided by the test standard deviation of scale score to obtain a standardized difference between the student score and the score a student with average ability would receive. This difference is then used to obtain the student standard score based on the normal distribution of student ability.

ATAC II OPERATIONAL FLOW



Student grades are used for several purposes: failure/attrition control, instructional effectiveness analysis, student performance ranking, and assignment of marks.

In order to be used for these purposes, the grades assigned to students must have two basic characteristics:

- Stable Criterion Measurement
  - All students of equal ability and performance should receive the same grade
  - Each different version of the same test should produce the same score for a given student
- Broad Discrimination Power
  - All tests should be constructed so that the number of tie scores is minimized (so as to minimize the number of students who are ranked equally)
  - Discrimination should be particularly strong at both ends of the ability scale, so that the poorest and the best students may be clearly identified.

The use of standard scores provides both of these characteristics.

The outputs from the test scoring function are a student answer card listing and an updated test item data file.

Student Grading and Academic Record Keeping

The grading process involves combining student scores on the daily tests and weekly test to obtain a weekly composite grade, ranking of students within a class, and generating a weekly student grade report and student academic history report.

During each week of instruction, from zero to five daily grades can be included in a student's record. Each daily grade is given a weight; the total of the daily grade weights must be 100. A weighted average of the daily grades is then computed to provide a composite daily grade.

Each students cumulative grade, rank and status is reviewed to determine whether or not he should be disenrolled from the school and transferred to other duty. Instructors can compare each students daily, weekly and cumulative grades to determine if there are any areas of unusual weakness for a student so that he can be given additional instruction in those areas. Daily grades for each student can also be compared to his performance on related questions in the weekly test (obtained from the student answer card listing) in order to determine if a student comprehends the subject material or merely recalls it on the daily test and forgets it when taking the weekly test. The comparison of daily grades and performance on related questions in the weekly test can also be used to identify the effects of student preparation time on his test performance since the student has more time to study and prepare for the weekly test than for the daily tests.

The weekly overall class performance is reviewed to identify changes in instructor effectiveness, the adequacy of course material covered during the week, and the adequacy of tests given during the week.

Test and Question Statistical Analysis

Test statistical analysis is performed primarily in order to evaluate instructional effectiveness. The basic measure of instructional effectiveness which is utilized by ATAC II is the difference between expected and actual test difficulty level. Actual test difficulty level is obtained from the weekly test results, and the expected test difficulty level is obtained from previous statistical data on the individual questions selected for the weekly test. The results are presented in an instructional effectiveness report. This report lists the number of students tested previously, number of students taking the current test, test reliability, number of questions, expected and actual test difficulty level, and the difference between the expected and actual difficulty level for each weekly test.

The primary purpose of the instructional effectiveness report is to compare expected and actual test difficulty levels. Whenever a difference

greater than $\pm$ 10 occurs a significant difference comment is printed in the interpretation column. The difference in expected and actual test difficulty level is a direct measure of instructional effectiveness provided test reliability is high. If a test is not reliable, the difference in expected and actual test difficulty level cannot be used as a measure of instructional effectiveness since the test does not provide a consistant measure of student academic ability.

Item analysis consists of statistical analysis of individual test items of weekly tests. These analyses are performed in order to monitor the quality and effectiveness of the test items and instruction. Two reports are generated for this purpose: the question analysis report and the item response table.

Statistical analysis of individual test items involves computing item difficulty level and discrimination index, and percentages of student responses for each item alternative. Item difficulty level is a direct measure of the item difficulty. It is the percentage of students exposed to the item who answer the item correctly. Multiple choice questions always allow a student who does not know the correct answer to guess; therefore, a guessing factor of one over the number of alternative choices on the test item is subtracted for each incorrect response to the item by a student. This gives an unbiased measure of the item difficulty. In military courses where the primary purpose is to train students for field operations rather than to academically segregate students, test items in general should have a difficulty level above 50 to be meaningful. Item difficulty levels in the high nineties indicate poor questions since almost everyone answers them correctly and they serve no purpose. There are exceptions, of course, often an easy question with a high difficulty level serves to remind students of basic and critical principles.

The other statistical measure of individual item effectiveness is discrimination index. In order to be effective, a test item must discriminate between good and poor students. The ability of an item to discriminate should be independent of the difficulty level of the item. The basic purpose of the

Marine Corps school is to train students to perform adequately in the field; therefore, the discrimination index should discriminate primarily between adequate and unsatisfactory students. The index which best satisfies these requirements (independent of difficulty level and discriminate between adequate and unsatisfactory students) is the biserial R statistic. The theoretical background information on and the derivation of this statistic is contained in the book Educational Measurements by Lindquist. The biserial R discrimination index is given below.

$$R = \frac{(\overline{X}p - \overline{X}q)}{\delta_t} \frac{p \cdot q}{y}$$

where: $\overline{X}_p$ = mean test scale score of students with correct response

$\overline{X}_q$ = mean test scale score of students with incorrect response

δ_t = standard deviation of test scale scores for all students

p = percentage of students with correct responses

q = percentage of students with incorrect responses

y = ordinate of normal distribution at cumulative area equal to p.

The percentage of student responses for each alternative answer on each item is also computed and listed in the item response table to provide further insight into the adequacy of a test item.

Statistical analysis of weekly tests involves computing percentile mean and range, expected and actual test difficulty level, test reliability, standard error of measurement, and scale score mean, standard deviation, mean standard error, median and range. The primary purpose of test statistics is to measure and monitor the consistency or reliability with which the test measures student academic ability. The difference between actual and expected test difficulty level, test reliability, and standard error of measurement provide direct measures of test consistency. The difference between expected and actual test difficulty level compares student performance on the test with historical

student performance on individual items in the test. Test reliability is computed using one of the standard educational statistics for test reliability given below.

$$r = \frac{n}{n-1}\left[\frac{N\Sigma X^2 - (\Sigma X)^2 - N\Sigma Y + N\Sigma Y + \Sigma Y^2}{N\Sigma X^2 - (\Sigma X)^2}\right]$$

where: r = test reliability

n = number of test questions

N = number of students taking the test

X = number of correct answers for each student

Y = number of correct answers for each question

The standard error of measurement separates the variation in test measurement from the variation in test scores and ability among different groups of students. It provides a measure of how accurately the test measures student academic ability.

The formula for standard error of measurement is given below.

$$s = \delta\sqrt{1 - r}$$

where: s = standard error of measurement

δ = scale score standard deviation

r = test reliability

The scale score mean, standard deviation, mean standard error, median and range provide the statistical parameters necessary to define the distribution of student scale scores on the test. These parameters will vary from test to test because of varying test difficulty levels and varying student ability from class to class. However, these parameters should stay fairly constant from week to week for the same class.

The question analysis report lists each item on the weekly test, the date the item was last used, the cumulative, current, and composite number of students exposes to the item, the cumulative, current and composite item difficulty level and discrimination index, and the test and item significance.

In addition, the question analysis report provides summary statistics for the weekly test. These statistics include percentile mean and range, expected and actual test difficulty, test reliability, standard error of measurement, number of students taking the test, and scale score mean, standard deviation, mean standard error, median and range.

This report is used primarily by the Test Control Group to monitor the quality and effectiveness of the test items. Difficulty level (percentage of students answering the question correctly) measures the relative difficulty of the item. A low difficulty level indicates the question may be too difficult, ambiguous or poorly worded. A high difficulty level indicates that a question is too simple and does not adequately test a student's ability. The difficulty level is adjusted for guessing by subtracting one over the number of choices for each wrong answer by a student.

The discrimination index is a measure of how well a question measures a student's academic ability. Questions with a large positive discriminator index distinguish well between students with greater and less ability. A large positive index indicates that most of the students with good academic records answered the question correctly while most of the students with poor academic records answered the question incorrectly. A discrimination index near zero indicates that the question is inadequate for distinguishing between good and poor students. This situation often occurs when a question is either extremely difficult or extremely easy so that nearly all or nearly none of the students answer the question incorrectly. A negative index indicates that more poor students than good students answer the question correctly. This can be caused by a question with a built in subtlety which causes problems for a good student but goes unnoticed by a poor student. It can also be caused by an extremely difficult question which many good students leave unanswered and many poor students answer by guessing. The cumulative difficulty level and discrimination index is compared to the current difficulty level and discrimination

index for each item to determine if there is a significant change from the cumulative in either. Test significance lists a comment (significant or check) whenever a significant change occurs. A change in difficulty level could indicate a change in instructional effectiveness. A change in the same direction (increase or decrease) in the difficulty level for several items indicates either a change in instructional effectiveness or a change in student motivation.

A change in the discrimination index without a change in difficulty level for an item may indicate a change in the level of guessing on the item. This can be checked by referring to the item response table to see if the percentage of blank responses has changed.

The weekly test summary statistics provide information on test and instructional effectiveness as well as the distribution of academic ability in the class. The difference between the actual and expected test difficulty is a direct measure of instructional effectiveness. Test reliability and standard error of measurement measure the consistency with which the test measures student academic ability.

Since student ability is normally distributed, the percentile mean and range is adequate to measure the distribution of percentile scores. Scale score standard deviation and range are used to measure the spread of academic ability in the class. The scale score mean, mean standard error, and median are used to measure the skewness of ability in the class. Normally the magnitude of the difference between the mean and median will be less than the value of the mean standard error. If the mean is significantly greater than the median, the class has more above average students than usual. If the mean is significantly less than the median, the class has more below average students.

The item response table supplements the question analysis report. It indicates the percentage of students responding to each alternative on each test item. The cumulative percentage of responses for each alternative is also listed for each item. Two astericks appear between the current and cumulative percentages for the correct response. An error response indicates multiple answers were punched on the student answer card.

The item response table is used primarily to determine the level of guessing by students. A small percentage of blank responses in conjunction with evenly distributed percentages of incorrect responses for all incorrect responses indicates a high level of guessing. The item response table is also used to determine if an item has a misleading alternative choice. This would be indicated by a high percentage of incorrect responses for a single alternative on an item.

Differences between current and cumulative response percentages can also be used as a measure of instructional effectiveness. If significant changes occur in the response percentages for several questions in a related area, the instructor may be improving or degrading the classroom explanation of this area. A change from evenly distributed incorrect responses to a high percentage of incorrect responses for a single alternative with no significant change in difficulty level for an item may indicate that the instructor is confusing the students in this area.

Item Selection and Test Preparation

Item selection involves selecting test items from the item file for a specific test, test printing, preparation of an answer key card, preparation of an item selection report, and file updating.

In order to assure that the instructional staff concentrates on presenting course content rather than preparing students for tests and to assure that

students cannot memorize test questions, a large data bank of test questions is maintained. The standard item selection algorithm uses number of exposures and date of last selection to select test items. The items are sorted first by date of last exposure, then within each date on number of exposures. The item with the fewest exposures in each date is selected sequentially starting with the oldest date through two weeks prior to the current date. This process is then repeated for the items with the next lowest number of exposures in each date until the total number of test questions has been selected. If enough questions with last exposures two weeks or more prior to the current date are not available, the two week date restriction is removed and items which were on the tests during the previous two weeks are selected.

This selection algorithm accomplishes several objectives. It assures that all questions in the data bank will be utilized, it utilizes new questions more frequently so that valid statistical information can be obtained as soon as possible, it assures that each test will have a different set of questions, and it minimizes the possibility of the same question appearing on tests twice within two weeks.

Provision is incorporated in ATAC II to override the standard item selection algorithm in order to prepare special tests. There are two basic methods of exception test selection. The first is to input a specific course identification number and curriculum week number and have the system produce a regular test for test curriculum week using the standard item selection algorithm. The other is to input number of items desired from specific curriculum modules or specific items to be selected. Special tests which are requested for regular classes are processed and files updated in the same manner as for standard tests. Special tests which are requested for other than regular classes such as for reservists are not processed or graded by ATAC II and the item files are not changed to reflect item use on the special test.

Each time an item is selected for a regular class test the number of exposures and date of last exposure for the item is updated in the item file.

After the items for the weekly test have been selected, ATAC II prints the test in the proper format for use by the students. It also prepares an answer key card which lists the correct answer for each item on the test for the instrcutors use in discussing the test questions after the students have completed the test. An item selection report which lists the items selected and their date of last use, is also generated in order to monitor the propor execution of the item selection algorithm.

Graduate Evaluation

The basic purpose of graduate evaluation is to provide a field operations feedback evaluation of the school curricula and training effectiveness. Graduate evaluation involves having graduates from terminal courses and supervisors of the graduates fill out questionnaires concerning the frequency with which the graduate performs tasks for which the courses prepared him, the graduate's ability in performing the tasks, and the graduate's opinion of the effectiveness of his training for the tasks. These questionnaires thus provide feedback information on the usefulness of the curricula and the effectiveness of the training.

Questionnaires are used because many of the tasks which the graduates perform do not have readily quantifiable measures of performance. There are two major problems with questionnaires. The first is the high incedence rate of errors due to personnel aversion to filling out questionnaires and due to key punch errors when converting to computer input cards. The second is the inability to use absolute questionnaire results because of the wide variation in personnel attitudes and value scales.

The first problem, high rate of input errors, is minimized by two edit programs. The first edits the identification fields of the questionnaire records, and the second edits the actual response fields of the questionnaire records.

Both edit programs print out error reports listing all records with errors. No record with an error is entered into the graduate evaluation data bank.

The second problem is alleviated by converting questionnaire responses to standard scores which measure the variation of the individual response from the historical average in terms of standard deviations from the mean.

In order to monitor the consistency and reliability of the questionnaire responses, a number of correlation coefficients between graduate and supervisor responses for the same course are computed. The classical statistical formula for correlation coefficient between two variables is utilized.

The results are presented in a graduate evaluation report which is used by school administrators to assist in monitoring the curricula to decide when changes are desirable and to evaluate long term trends in instructional effectiveness.

The graduate evaluation report lists the scale score for questionnaire responses of frequency of task performance by the graduate and his supervisor, ability of the graduate in task performance by the graduate and his superior, course value in task performance by the graduate, and six different response correlations for each task.

The graduate evaluation report is used by the school administrative staff as a measure of training effectiveness. High frequency of task performance scores indicate the relative need for a course. Task performance ability and course value scores are a direct measure of course effectiveness.

The six correlations (graduate to supervisor frequency, graduate to supervisor ability, graduate frequency to ability, supervisor frequency to ability, graduate course value to graduate ability, and graduate course value to supervisor ability) provide both a measure of reliability of questionnaire responses and additional measures of training effectiveness. High graduate to supervisor frequency and graduate to supervisor ability correlations indicate that the questionnaire responses are consistent and reliable. A high frequency to ability

correlation would indicate that the graduate benefits mostly from on-the-job experience. Comparison of graduate and supervisor frequency to ability correlations provides a measure of questionnaire response consistency.

Course value to ability correlations provide a measure of training effectiveness. High course value to ability correlations indicate that ability is due to training received in the course. Comparison of graduate and supervisor course value to ability correlations provides a measure of questionnaire response consistency.

III. ATAC II as an Improved Technique

ATAC II represents a significant advance in the state-of-the-art of academic testing. It provides a number of advantages over existing methods of performing academic administrative functions including speed of scoring, grading, and updating student academic records, increased accuracy of scoring and grading, significant cost savings in school operating costs, increased statistical validity of grading and attrition control, immediate feedback and corrective action for deficiencies in student performance and instructional effectiveness, a large unbiased selection of test questions, and rapid updating of the test question data bank.

ATAC II can score tests, grade students, and update student academic records for 2,000 students in less than two hours. All of the functions of ATAC II, including statistical analysis, file updating, and test item data bank updating can be accomplished for 2,000 students in less than 15 hours per week,

Scoring and grading is more accurate than existing methods for two basic reasons. The first is that many errors occurring in the manual processing of tests and computing grades are eliminated through automatic scoring and grading. The second is that ATAC II contains numerous editing programs which edit all information which is input to the system. Error reports are produced so that the information can be corrected and re-input to the system before it is processed.

ATAC II significantly reduces school operating costs. The Communication-Electronics School started using an interim ATAC system with less capability than the current system in 1964. In a six-year period, the C-E School has documented savings in operating costs of over 2.5 million dollars. The school administrative staff has been reduced from 67 to 13. Annual operating costs per student have been reduced from $103 to

ATAC II makes the use of statistical data from previous students feasible for grading and attrition control of current students. This provides a much larger sample size for statistical grading and attrition control. This significantly increases the statistical validity of the grading and attrition control process. Typically the sample size is one to two orders of magnitude larger than a single class.

Because ATAC II provides all output reports during the week following a week of course instruction and testing, deficiencies in student performance and instructional effectiveness can be determined and corrected immediately. Student deficiencies can be corrected through tutoring, setback and course repeat, or academic dropping and reassignment. Instructional deficiencies can be corrected through additional instructor guidance or training, changes in course material, or changes in test material.

ATAC II currently has a test question data bank of 50,000 test questions. This forces the instructional staff to concentrate on presenting course content rather than preparing students for tests. It also prevents students from memorizing question answers or obtaining prior information on test question content. A large data bank of test questions is not feasible using manual methods of test preparation.

Todays rapidly changing technology and operational equipment makes it imperative that course material and test questions be continuously updated in order to properly prepare the student for field operations. ATAC II permits weekly updating of the test question data bank.

In summary, large volumes of students, rapidly changing technology, and reduced operating budgets and personnel availability, make automated testing and administrative functions a necessity for military methods. ATAC II represents a major advance in automating these functions.

August 1971

THE USE OF LOGIC TREES IN MILITARY PERFORMANCE TESTING

by

Raymond L. Erickson

Presentation to
1971 Military Testing
Association Conference
Washington, D.C. September 1971

THE USE OF LOGIC TREES IN MILITARY PERFORMANCE TESTING

Raymond L. Erickson

INTRODUCTION

The creation of an effective performance test is normally the culmination of exhaustive analysis and planning. Innumerable methods have been devised to ease the test design effort, and all doubtlessly contribute to the development of realistic, objective and comprehensive tests which are capable of providing feedback specific enough to control instructional quality. Despite the plethora of test design techniques, the United States Army Adjutant General School has found one analytical tool to be far superior to all others in the construction of valid instruments for the measurement of student achievement. At the Adjutant General School, the first significant step in the creation of a performance test is the construction of a Logic Tree.

DEFINITION

The Logic Tree is formally defined as a schematic representation of a mental decision making process and the actions that result from such decisions. Quite simply, the Logic Tree is a decisional flow chart. The Logic Tree is decisional in that it graphically depicts each of the decisions which must be made in the performance of the task being analyzed, from the initiation to the completion of that task. The Logic Tree is a flow chart since each of the decisions in the performance of the selected task is placed in its most logical sequence. Consequently, to Logic Tree a task is merely to list all the decisions which must be made in the performance of that task, from start to finish, in their mos logical order.

LOGIC TREE CONSTRUCTION

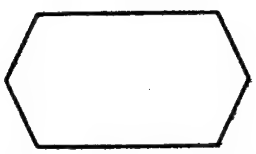
The structure of a Logic Tree can be best described by dividing the analytical tool into its two component elements: 1) the cover sheet; and, 2) the decisional schematic.

The cover sheet of a Logic Tree performs the same function as the table of contents in a textbook. The properly completed cover sheet should disclose information sufficient enough to inform the reader whether or not this is the Logic Tree he desires to read. Figure number 1 demonstrates the content and format of a cover sheet. At the top of the cover sheet is found the necessary statements of a Training Objective. The Training Objective will identify the task being analyzed and the conditions (givens) under which the task is to be performed.

The conditions segment of the Training Objective must also include the cue or the stimulus which causes the task to be initiated. Also to be noted on the cover sheet is the listing of the Source Data. Such a listing ensures that the reader can check upon the accuracy of the Logic Tree's content.

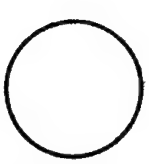
The decisional schematic of the Logic Tree is merely a collection of blocks and directional lines. Figure number 2 depicts the structure of a Logic Tree schematic. As noted in that Figure, the Logic Tree is composed of four types of symbolic blocks, each having a separate function and meaning.

The oval represents either the beginning or the ending of the task being analyzed. It is worthy of comment that the Start or beginning of the task block incorporates by definition the "conditions/cues" statement of the cover sheet. By means of this rule of construction, the beginning of the task and all the necessary conditions and cues are firmly established.



All the decisions to be made in the selected task will be found within the confines of a hexagon. From Figure number 2, it is readily apparent that each decision is written in the form of a question answerable only by "yes" or "no". Consequently, the decisional block must have two exit, and only two exit, paths; one for the "yes" and the other for the "no" decision. Any information required to answer a decisional question represents knowledge necessary to the performance of the task.

Rectangular blocks contain information or instructions helpful to the continuance or to the completion of the task being analyzed. Since no decisions are to be found within a rectangle, such blocks will have but one path of exit.



Finally, the circle represents an exit or a jump to another portion of the Logic Tree. Through the use of this block, the reader may jump over any unnecessary steps in the performance of the task to that location in the Logic Tree where the task procedure again becomes relevant to him. Obviously, the circle has no exit path since it contains written instructions as to where in the schematic the reader should proceed.

All that remains of the Logic Tree schematic to be described are the connecting lines and the directional arrows. These elements do nothing more than graphically depict the various decisional paths open to the

reader. By means of such guidelines, the possibility of the reader departing from the logical sequence of the decisions is effectively precluded.

RULES OF LOGIC TREE DESIGN

The rules of Logic Tree design are intended to ease the effort of Logic Tree construction and to enhance the clarity and usefulness of the schematic itself. Common sense in the design of Logic Trees dictates that the more elementary the schematic, the more beneficial the Logic Tree will be to the student, the instructor and the training analyst.

At this stage, it demands little comment that Logic Trees can be constructed only to depict tasks and never to graphically describe a subject or general topic. The task can be mental, physical or a combination of the two, but it must have a definite starting point, a definite ending, and must be performed because of certain conditions or circumstances which are found to exist.

As well, the Logic Tree schematic should be so designed so as not to refer the user to any regulation or outside informational source, if at all possible. The Logic Tree is intended to supplant such informational sources.

Equally important, exceptions to the general rule and decisions or actions common to all major decisional branches should be placed near the beginning of the Logic Tree, thus eliminating the needless repetition of such blocks in each of the different trunk-lines of the Tree. Also with reference to exceptions to the general rule, the user will immediately find the exception and if it is relevant to him, can follow the decisional path pertaining to that exception without first traversing the decisions of the entire Tree.

Finally, by keeping the "Yes" and "No" beside the decision block at the very beginning of the exit line, by refusing to cross one directional line over another, by using sufficient directional arrows, and by numbering the symbolic blocks from left to right and from the top to the bottom of the page, the clarity and simplicity of the Logic Tree will be greatly enhanced.

THE LOGIC TREE AND MILITARY PERFORMANCE TESTING

The properly constructed Logic Tree has numerous uses in the process of course design, and obviously, each of these uses contributes somewhat to the development of effective performance testing devices. However, due to the limited scope of this paper, only those uses of the Logic Tree which are essential to test design will be discussed.

In accordance with the Systems Training concept as established by CONARC Regulation 350-100-1, the training analyst must minutely analyze the various tasks which he has selected for training. Such a Task Analysis is undertaken to disclose all the decisions which are essential to the performance of the task being analyzed.

At the Adjutant General School, the Logic Tree functions as the Task Analysis step in the Systems Engineering process primarily because it compels disciplined, logical thinking. Through the use of the Logic Tree, the analyst can readily determine the exact nature of the task and what skills and knowledges the student must master in order to perform the selected task.

The Logic Tree has inherent advantages over other Task Analysis methods namely because it requires the analyst to express each mental element of the task in the form of a question. By further requiring each question to be answered by a "yes" or a "no", the analyst is compelled into considering all possibilities and he can consequently uncover aspects of the task which would not have been so apparent under a less methodical approach.

Once the Logic Tree has been prepared, the creation of the performance test itself becomes vastly simplified. Since the Logic Tree graphically depicts every decision within the task and every end result of that task, the analyst need only pick the appropriate decisions and results he considers worthy of testing. By drawing a line through the chain of decisions so tested, the analyst need not worry about over-testing any particular variation of the task. In one procedure then, the training analyst has placed all the elements of the task before his evaluatory eye, and has ensured a comprehensive and valid examination. Less methodical approaches function chiefly as a means of documenting the analysis which has been undertaken, while the Logic Tree not only documents such analysis but constitutes the analytical tool itself.

To demonstrate the use of a Logic Tree in the development of a performance test, consider again the Logic Tree schematic shown at Figure number 2. This Logic Tree graphically represents a block of instruction presented in the Adjutant General School's Instructor Training Course.

It is the policy of the Adjutant General School that every instructor must not only command the training platform with confidence, but must also be able to continuously evaluate the effectiveness of his instruction. In pursuance of this policy, each instructor is given computer print-outs covering the students' performance on that instructor's examination. In particular, the instructor must be able to evaluate the students' responses to each question/problem included in the examination as disclosed by the Item Analysis print-out. Figure number 3 depicts a typical Item Analysis print-out. Note on that figure that items number 4 and 8 have miss rates in excess of ten percent, i.e. more than ten percent of the students in the class failed to answer those questions correctly. At the Adjutant

General School such miss rates are considered unacceptable and demand the immediate attention of the instructor.

When confronted by an excessive miss rate, the instructor can consider that one of two contingencies caused the problem: 1) his instruction was unclear and ineffective; or, 2) the test instrument was misleading and invalid. Since the instructor has in his possession a copy of the answer key, a copy of the test instrument and his own personal lesson plan, each of the causative factors can be explored. For those who would also add the possibility that the excessive miss rate was due solely to student lack of effort, the countervailing argument seems far too strong to allow such consideration. The School allows up to ten percent of the class to miss a question due to such causative factors of poor student motivation as headaches, poor scheduling of an examination, Spring Fever and whatever else would distract a student's attention from successful performance on an examination. Should a higher percentage of students be so distracted by such personal causative factors, then a re-evaluation of the entire course would seem warranted.

In the Instructor Training Course, the student is first presented the foregoing instructional information and is then presented with a practice version of the performance examination in order that he might practice his skill of self-evaluation. This practice version of the examination is shown at Figure number 4. Note that the examination first places the student in the position of an instructor of the administrative review of Disposition Forms, and that the Instructor Training student is given all the tools normally possessed by an incumbent instructor in the Adjutant General School. The student is then required to evaluate the Item Analysis print-out as found on the practice test instrument and determine if the excessive miss rates were due to ineffective instruction or due to errors in the design of the administrative review test which they administered to a hypothetical class.

For example, item number 4 on the TEST RESULT PRINT OUT discloses a miss rate of 13%. Such a miss rate is unacceptable and demands immediate action on the part of the instructor. The instructor (Instructor Training student) must then check the ADMIN REVIEW ANSWER SHEET KEY to discover that item number 4 dealt with the FROM block on the Disposition Form. Next the Instructor Training student must check the ADMIN REVIEW TEST INSTRUMENT to determine if the unacceptable miss rates were caused by some error in the design of the test such as smudged or illegible printing. By merely looking at the FROM block on the ADMIN REVIEW TEST INSTRUMENT, the instructor can readily perceive that the printing is legible and if the hypothetical student who was administered the ADMIN REVIEW TEST INSTRUMENT had known the subject matter of administrative review, he would have recognized that the Disposition Form came from the Adjutant General and not from the Administrative Services Division. Obviously, if the test instrument is not defective, then ineffective instruction must have caused the excessive miss rate and a review of that portion of the lesson

plan covering the FROM block is in order. The Instructor Training student must then examine the abbreviated LESSON PLAN found on the test to determine which paragraph of that LESSON PLAN need be reviewed. In this particular case, paragraph number 4.d. needed review and the Instructor Training student would transfer that information to the Answer Sheet and mark the answer block opposite paragraph 4.d. in Column A (answer block number 4). Had the excessive miss rate been caused by an error in the test design, then block number 17 on the answer sheet would have been marked. Item number 12 on the TEST RESULT PRINT OUT concerns itself with a test design error, namely a missprinting in the Inclosures portion of the Disposition Form. Obviously, the test is not completed until the cause for each unacceptable miss rate has been isolated and identified. Following this practice test, the Instructor Training student is then required to complete a graded examination which takes exactly the same form and requires the execution of exactly the same task.

In relating this particular examination back to its parent Logic Tree, it becomes readily apparent that the Logic Tree depicted each of the three end results which were finally adjudged to be worthy of testing: 1) the training was effective as disclosed by an acceptable miss rate; 2) the unacceptable miss rate was due to ineffectual instruction; and, 3) the excessive miss rate was generated by faulty test design.

To denote that a particular end result had been tested, a colored line was drawn through the appropriate blocks in the Logic Tree. Each of the possible end results was tested at least once. For example, the first end result above was tested in item number 1 on the TEST RESULT PRINT OUT. That particular item tested blocks 1, 3, 9, 10, and 4 on the Logic Tree and a colored line denoted that decisional path. The second end result above was presented in item number 4 on the TEST RESULT PRINT OUT. Consequently, item number 4 tested blocks 1, 3, 9, 14, 18, 23, 19, 20, 24 and 21. A different colored line was drawn through that set of blocks. Finally, the last possible end result was presented to the student by item number 12 on the TEST RESULT PRINT OUT. That item tested blocks 1, 3, 9, 14, 18, 23, 19, 20, 16, 11 and 5. A third color was used to indicate this final decisional path. Through this technique of first discovering every possible decision within the task of evaluating an Item Analysis Print Out and every possible end result, the task in its entirety could be presented to the student both during the instruction and during the examination.

CONCLUSION

The Adjutant General School does not consider the Logic Tree as a panacea for all training problems, but consistently this analytical device has proven itself to be an efficient and valuable tool in the preparation of realistic, objective and comprehensive performance tests which are also capable of providing specific feedback information.

Figure No. 1

LOGIC TREE FOR:

EVALUATION OF TEST RESULT PRINTOUTS

Task: To utilize an Item Analysis test result print-out in order to eliminate deficiencies in instruction and test design.

Conditions/Cues: Receipt of an Item Analysis test result print-out, and access to the appropriate answer sheet key, Test Instrument and Lesson Plan.

Source Data

USAAGS Reg 350-2, dtd 7 May 69 W/C1
USAAGS Reg 350-100, dtd 1 Jul 68 W/C1

SUPERSEDES: NA

25 Aug 71

FIGURE NO. 2

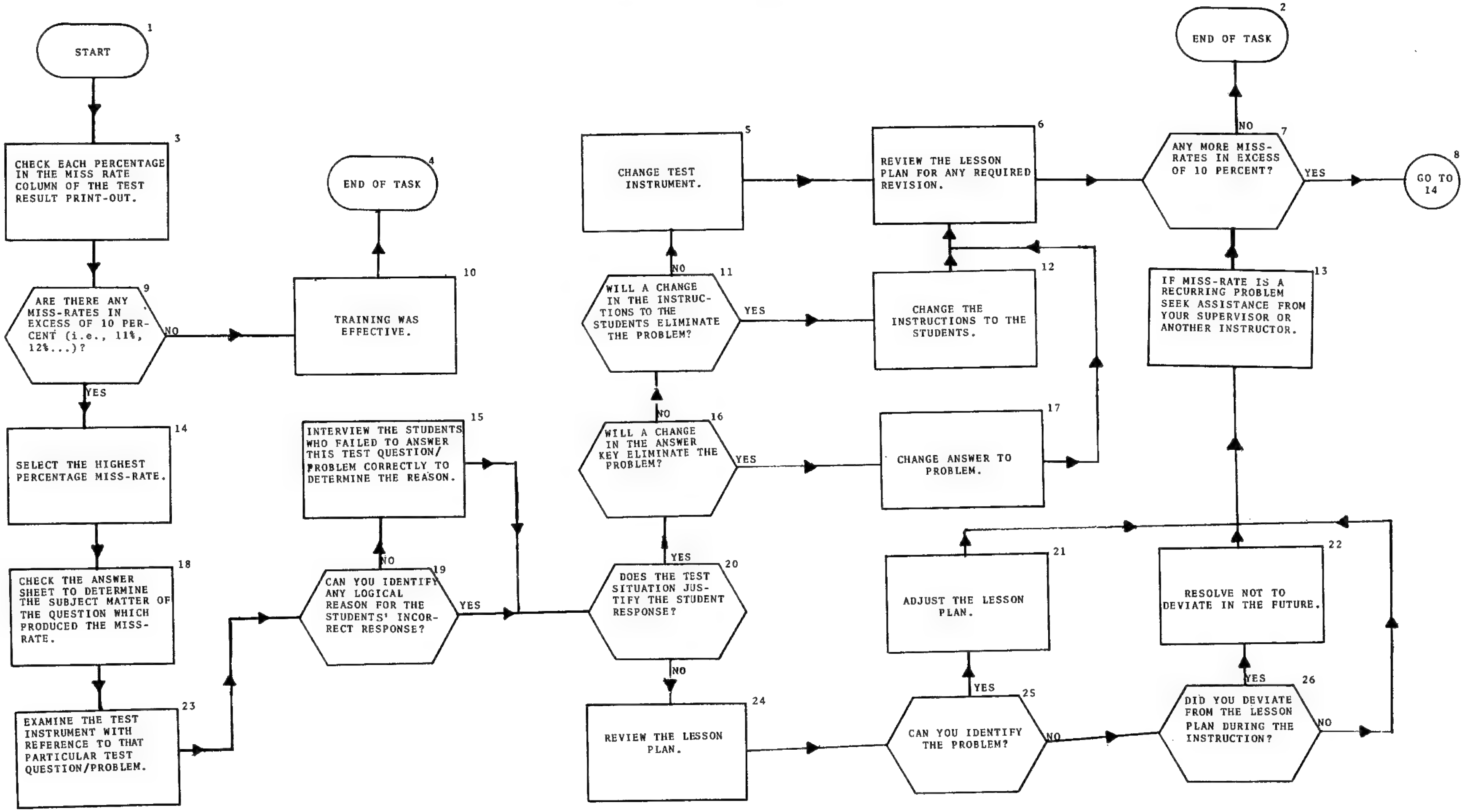

START 1
CHECK EACH PERCENTAGE IN THE MISS RATE COLUMN OF THE TEST RESULT PRINT-OUT. 3
ARE THERE ANY MISS-RATES IN EXCESS OF 10 PER-CENT (i.e., 11%, 12%...)? 9
NO
YES
TRAINING WAS EFFECTIVE. 10
END OF TASK 4
SELECT THE HIGHEST PERCENTAGE MISS-RATE. 14
CHECK THE ANSWER SHEET TO DETERMINE THE SUBJECT MATTER OF THE QUESTION WHICH PRODUCED THE MISS-RATE. 18
EXAMINE THE TEST INSTRUMENT WITH REFERENCE TO THAT PARTICULAR TEST QUESTION/PROBLEM. 23
CAN YOU IDENTIFY ANY LOGICAL REASON FOR THE STUDENTS' INCORRECT RESPONSE? 19
NO
YES
INTERVIEW THE STUDENTS WHO FAILED TO ANSWER THIS TEST QUESTION/PROBLEM CORRECTLY TO DETERMINE THE REASON. 15
DOES THE TEST SITUATION JUSTIFY THE STUDENT RESPONSE? 20
YES
NO
WILL A CHANGE IN THE ANSWER KEY ELIMINATE THE PROBLEM? 16
NO
YES
WILL A CHANGE IN THE INSTRUCTIONS TO THE STUDENTS ELIMINATE THE PROBLEM? 11
NO
YES
CHANGE TEST INSTRUMENT. 5
CHANGE THE INSTRUCTIONS TO THE STUDENTS. 12
CHANGE ANSWER TO PROBLEM. 17
REVIEW THE LESSON PLAN FOR ANY REQUIRED REVISION. 6
ANY MORE MISS-RATES IN EXCESS OF 10 PERCENT? 7
NO
YES
END OF TASK 2
GO TO 14 8
IF MISS-RATE IS A RECURRING PROBLEM SEEK ASSISTANCE FROM YOUR SUPERVISOR OR ANOTHER INSTRUCTOR. 13
REVIEW THE LESSON PLAN. 24
CAN YOU IDENTIFY THE PROBLEM? 25
YES
NO
ADJUST THE LESSON PLAN. 21
DID YOU DEVIATE FROM THE LESSON PLAN DURING THE INSTRUCTION? 26
YES
NO
RESOLVE NOT TO DEVIATE IN THE FUTURE. 22

FIGURE NO. 3

ITEM ANALYSIS
DATE - 19AUG70

CARD NO. - 5 OF 5

COURSE - AGOBC
CLASS - 71-02

TEST - AD SVCS
SITUATION - AD REVIEW

CARD TYPE - B
RAW POSS - 25

ADMINISTRATIVE REVIEW CARD 5-5

QUES	PUN	UNPUN	MISS RATE
1	4	47*	7%
2	4	47*	7%
3	49*	2	3%
4	35*	16	31%
5	5	46*	9%
6	49*	2	3%
7	1	50*	1%
8	8	43*	15%
9	4	47*	7%
10	46*	5	9%
11	2	49*	3%
12	5	46*	9%
13	5	46*	9%
14	47*	4	7%

The * indicates a correct student response.

ADMIN REVIEW - DISPOSITION FORM

1.	Office Symbol	/1/
2.	Subject	/2/
3.	TO	/3/
4.	FROM	/4/
5.	DATE/Originator	/5/
6.	Paragraph 1	/6/
7.	Paragraph 2	/7/
8.	Paragraph 3	/8/
9.	Paragraph 4	/9/
10.	Authority Line	/10/
11.	Signature Block	/11/
12.	Enclosures	/12/
13.	Distribution	/13/
14.	Copies Furnished	/14/
15.	Page Number	/15/
16.	Number of Copies	/16/

ITEM ANALYSIS
DATE - 19OCT70

CARD NO. - 5 OF 5

COURSE - AGOBC TEST - AD SVCS C CARD TYPE - B
CLASS - 71-8 SITUATION - AD REVIEW RAW POSS 32

ADMINISTRATIVE REVIEW CARD 5-5

QUES	PUN	UNPUN	MISS RATE
1	4	48*	7%
2	3	49*	5%
3	1	51*	1%
4	45*	7	13%
5	46*	6	11%
6	47*	5	9%
7	49*	3	5%
8		52*	
9		52*	
10	1	51*	1%
11	42*	10	19%
12	13	39*	25%
13		52*	
14		52*	
15		52*	
16	37*	15	28%

GENERAL SITUATION:

You are an instructor in the United States Army Adjutant General School and you have just received the test result print outs for the first class which you have instructed on the Administrative Review of Disposition Forms. You decide to evaluate the item analysis portion of the test results in order to determine which areas of the Lesson Plan should be reviewed. In addition, you decide to review the test instrument based on the information contained in the item analysis in order to determine if any corrections in the test are required.

REQUIREMENT:

Based on the information contained on this test supplement, place a check mark in the numbered block on the answer sheet under Column A to denote which paragraphs of the Lesson Plan should be reviewed, and under Column B to denote an error in the construction of the test instrument which contributed to the unacceptable miss rate.

ANSWER SHEET

Paragraphs of the Lesson Plan.	Column A Lesson Plan should be reviewed.	Column B Test instrument contains an error.
4.		
a.	/1/	/14/
b.	/2/	/15/
c.	/3/	/16/
d.	/4/	/17/
e.	/5/	/18/
5.		
a.	/6/	/19/
b.	/7/	/20/
c.	/8/	/21/
d.	/9/	/22/
6.		
a.	/10/	/23/
b.	/11/	/24/
c.	/12/	/25/
d.	/13/	/26/

LESSON PLAN

(Abbreviated for Test Purposes)

ADMINISTRATIVE REVIEW OF THE DISPOSITION FORM

A. Heading of the Disposition Form
B. Body of the Disposition Form.
C. Closing of the Disposition Form.

SECTION I-INTRODUCTION

1. Attention.

2. Motivation.

3. Objectives.

SECTION II-BODY

4. First Mail Teaching Point. Heading of the Disposition Form.

   a. Office Symbol or Reference.
   b. Subject Block.
   c. TO Address.
   d. FROM Address.
   e. DATE/ORIGINATOR.

5. Second Main Teaching Point. Body of the Disposition Form.

   a. Detection of errors in spelling, grammar and punctuation in body of DF.
   b. Paragraph numbering.
   c. Use and lettering of subparagraphs.
   d. The Modified block style format.

6. Third Main Teaching Point. Closing of the Disposition Form.

   a. Use of the Authority Line.
   b. Format and use of the Signature Block.
   c. Identification of Enclosures.
   d. Use of Copies Furnished and the preparation of copies.
   e. Continuation Pages and page numbering.

SECTION III-CONCLUSION

7. Questions.

8. Summary.

9. Closing.

DISPOSITION FORM
For use of this form, see AR 340-15; the proponent agency is The Adjutant General's Office.

REFERENCE OR OFFICE	SUBJECT		
AJJAG	Sponsor for New Officer		
TO G3	FROM ASD	DATE 15 May 1974 2LT Campbell/kw/2935	CMT 1

1. We have just received the attached orders assigning CPT John C. Scott to the division. Since CPT Scott is tentative scheduled to be assigned to your section,I, request you designate a sponsor for him.

2. In accordance with General Johns' policy, a letter and a division welcome packet will be forwarded to CPT Scott not later than 20 May 1974 and an information copy of the letter will be furnished this office.

1 Encl

W. W. GEMMILL
LTC, AGC
Adjitant General

DA Form 2496 1 Feb 62 — Replaces DD FORM 96, existing supplies of which will be issued and used until 1 Feb 63 unless sooner exhausted.

DIGITAL COMPUTER SIMULATION: A TOOL FOR PSYCHOLOGISTS

William A. Sands

Naval Personnel Research and Development Laboratory[1]
Washington, D. C.

INTRODUCTION

Simulation is basically a research approach to problem-solving. A model of some real world system or situation is constructed and then experiments are performed on the model to provide information about the actual system.

Johnson (1968) describes a model as: "A logically connected set of rules that abstract selected characteristics of some phenomena or system." Rhodes (1970) indicates that: "A model can be thought of, in simplest terms, as a function set processing an input set, producing an output set, corresponding to processes, inputs, and outcomes in reality."

Martin (1968) defines computer simulation as: "A logical-mathematical representation of a concept, system, or operation programmed for use on a high-speed computer."

TYPES OF MODELS

Any taxonomy of models is, to some extent, arbitrary. One method of classifying models focuses on the problem the model is designed to address: for example, inventory models, replacement models, and transportation models.

The following four ways of contrasting types of models appear useful for illustrating the diverse nature of models: (1) physical vs. mathematical models; (2) analog vs. digital models; (3) deterministic vs. stochastic models; and, (4) bulk flow vs. entity models.

---

1 The views expressed herein are those of the author and do not necessarily reflect those of the Navy Department.

Physical vs. Mathematical Models

Photographs, globes, a child's doll and scale models of jet aircraft are examples of physical models. Mathematical models, on the other hand, symbolize the real world referent with one or more equations. The advent of the high-speed, large memory computer made possible the development and utilization of mathematical models of complex systems (Hatch, 1969b). The United States Marine Corps Computer-Based Recruit Assignment Model (COBRA) is an example of a mathematical model of a complex personnel system (Hatch, 1969a).

Analog vs. Digital Models

An analog model employs physical magnitudes to represent numbers. For example, a slide rule uses the physical magnitude of length to represent the logarithm of numbers. In contrast, a digital model employs a series of digits to represent numbers. A desk calculator is an example of a digital device.

Deterministic vs. Stochastic Models

A deterministic model is one in which the output data are solely a function of the input data. Repeated applications of the model using the same input data yield identical output data.

The Cost of Attaining Personnel Requirements (CAPER) Model, developed by the Navy, is a deterministic model. The CAPER Model provides an optimal recruiting-selection strategy which minimizes the estimated total cost of recruiting, selecting, inducting, and training a sufficient number of persons to meet a specified quota of satisfactory personnel (Sands, 1970, 1971a, 1971c). Given the same set of costs and empirical frequency distributions, the CAPER Model will generate the same optimal recruiting-selection strategy and associated cost estimates.

Some authors (e.g., Bartholomew, 1967; Niehl and Sorenson, 1968) contend that the simplicity of analytic or deterministic models often makes them inadequate for studying personnel systems. They maintain that many of the input parameters which are treated as constants by deterministic models are not fixed, and should be viewed in probabilistic terms using stochastic models.

Hatch (1971) states that:

> "The interest in pure simulation models, as opposed to analytical models or simulation models with inbedded optimization, may be explained, in part, by the evolution of increasingly complex organizational structures and personnel policies, conditions which render impractical application of purely analytical models for detailed structure and policy assessment purposes. A simulation

> model is a procedural model in which the relevant processes and decisions are simulated by fairly conventional heuristic and logical procedures. As a result of their simplicity, simulation models can accommodate greater discrimination than can formal mathematical models. Simulation systems can therefore be created for a wide variety of problems and a high degree of discrimination. Pure simulators, sometimes called 'if, then' models, permit much greater penetration of their internal logic by mathematically unsophisticated users; consequently, such models enjoy considerable popularity for practical applications to manpower planning."

Analytic techniques which would handle complex personnel systems have been proposed (Boldt, 1962) but have not proved economical (Niehl and Sorenson, 1968).

In a stochastic model, one or more of the variables is considered to be random. The term "random," as used herein, does not indicate without rhyme or reason. Rather, the stochastic aspects of the system are treated in terms of random variables with specified probability distributions.

Gordon (1969) points out the impact of incorporating stochastic variables into a model:

> "Because of the interrelations between the activities of a system, the introduction of a stochastic variable into a system becomes reflected throughout the system. Most, if not all, of the quantities of interest in measuring the system performance then show random flutuations."

The SHIP II Model, developed by the Navy, is an example of a stochastic simulation model. This model is designed to realistically portray the numerous dynamic and complex interrelationships among manning requirements, equipment maintenance policies, and task requirements. SHIP II is event-oriented, producing random samples of events conforming to empirically derived probability distributions (Schwartz, 1971; Schwartz, *et al.*, 1970).

Bulk Flow vs. Entity Models

Another useful dimension for describing models is based upon the way in which individual entities are treated. In bulk flow models, persons (entities) with similar characteristics (e.g., time-in-service) are treated together as a group; i.e., the individuals lose their identities, (Johnson, 1971). The Career-Noncareer Model developed by the Army is an example of a bulk flow model. This model was designed for evaluating alternative personnel policies concerning training input, reassignment, manning levels, and manpower utilization (McMullen, 1969, 1970).

Entity models, on the other hand, represent explicitly and keep track of individuals in the system. The status of each person (entity) is known throughout the simulation experiment.

The United States Navy Computer-Assisted Recruit Assignment Model (COMPASS II) is an example of an entity model. According to the user's manual (Decision Systems Associates, Inc., 1969), COMPASS II was designed to meet the following system objectives:

(1) "Honor Navy enlisted procurement guarantees;

(2) Maximize quota accommodation through consideration of total Navy training opportunities and the total pool of recruits available for assignment at all Navy Training Centers;

(3) Minimize transportation costs associated with travel to advanced training and general duty locations;

(4) Maximize adherence to recruit preferences and interviewer recommendation policies;

(5) Within the constraints imposed above, maximize the probability of success of each school-assigned recruit in the school to which he is assigned; and,

(6) Assign lower mental standard personnel under separate Bureau of Naval Personnel distribution specifications."

PROBABILITY DISTRIBUTIONS

The random variation of the system portrayed by stochastic models can be simulated if the population of interest can be represented by a known probability distribution. Monte Carlo methods for simulating many probability distributions have been developed. Procedures for simulating the continuous and discrete probability distributions listed below are explained elsewhere (Naylor, et al., 1966).

Continuous Probability Distributions

1. Uniform distribution
2. Exponential distribution
3. Gamma distribution
4. Normal distribution
5. Multivariate normal distribution
6. Lognormal distribution

Discrete Probability Distributions

1. Geometric distribution
2. Negative binomial distribution
3. Binomial distribution
4. Hypergeometric distribution
5. Poisson distribution
6. Empirical discrete distribution

As Boldt (1965) points out, a procedure for simulating mixed normal and uniform statistical distributions is necessary in some cases. For example, when the Armed Forces Qualification Test (AFQT) is included in a simulation study along with the Army Classification Battery (ACB), the uniform distribution is introduced because the AFQT, a percentile measure, is uniformly distributed.

APPLICATIONS OF SIMULATION IN PSYCHOLOGY

Simulation techniques have been employed in the investigation of a variety of interesting psychological and psychometric problems. Niehl and Sorenson (1968) state that:

> "Relationships between requirements, assignment procedures, input from outside the system (for example, from external procurement or from training facilities), quantity and quality of personnel information made available to the system, attrition of the system, and measures of system effectiveness have been successfully studied through computer simulation."

Test Selection Methods

Harris (1967) employed multivariate normal score simulation to compare two methods of test selection: (a) differential prediction battery (Horst, 1954); and (b) multiple absolute prediction battery (Horst, 1955). The study was carried out in four stages: (1) selection of tests to be included in each type of predictive battery; (2) simulation of scores for the tests selected into each battery; (3) optimal assignment of the computer-generated persons or entities; and (4) evaluation of the results. The criterion of test battery effectiveness was the average expected performance of personnel optimally assigned to jobs on the basis of their simulated scores on the two types of test batteries. He found that the differential test battery resulted in a more effective assignment of personnel than the multiple absolute test battery.

Aptitude Area Scores vs. Full Regression Equations

The Army Classification Battery (ACB) is a collection of eleven tests designed to forecast performance in different job areas. In order to simplify computational problems, the Army operating system was basing predictions on composites of two tests, called aptitude area scores. Sorenson (1965a, 1965b) employed simulation techniques to assess the loss in assignment effectiveness resulting from the use of the aptitude area scores rather than the full eleven test regression equations. Performance estimates based upon the abbreviated and full set of tests were used to optimally assign computer-generated persons into eight job areas in such a way that the prescribed quotas were met. He found that, in comparison to the abbreviated composite, the gain over random assignment was roughly doubled by the use of the full eleven test regression equations.

Comparison of Classification Strategies

Alf and Wolfe (1968) used computer simulation to assess the efficacy of nine alternative personnel assignment strategies in terms of seven criteria of assignment effectiveness. The enlisted men in the sample were given simulated assignments to a Navy Class "A" school or the fleet nine separate times, one assignment for each of the alternative allocation policies. Identical school quota constraints were employed for each of the nine assignment strategies listed below:

(1) Actual assignment by the Naval Training Center;
(2) Random assignment;
(3) Maximize the sum of the scores on the operational school selection composite;
(4) Maximize estimated final grade average of school assignees;
(5) Maximize the average probability of school graduation for all school assignees;
(6) Minimize the average training cost, excluding pay and allowances, for all school assignees;
(7) Minimize the average training cost, including pay and allowances, for all school assignees;
(8) Minimize the manning level shortages;
(9) Maximize the criticality index (explained below).

The seven criteria of assignment policy effectiveness were:

(1) Aptitude school selection test scores
(2) Estimated final grade average
(3) Probability of school success
(4) Training cost, excluding pay and allowances
(5) Training cost, including pay and allowances
(6) Manning level, or the ratio of on-board strength to Navy needs
(7) Criticality index, reflecting both the number of personnel required in each category and the degree to which they were needed for the effective functioning of the Navy.

As expected, the authors found that the average payoff was always best for the assignment policy directed towards that criterion. Also, as expected, actual assignment and random assignment policies were not best on any of the criteria, since these strategies are not maximization policies in the mathematical sense. The actual assignments were far superior to the random assignments on all seven criteria. The best overall assignment policy was strategy #5, aimed at maximizing the probability of school success. This strategy came close to being the best on all seven criteria and also resulted in relatively homogeneous talent groups being assigned to the various schools.

Influence of Metric Changes on Assignment Algorithms

Simulation techniques were employed by Sorenson (1965a, 1966) to study the influence of metric changes in test scores employed by linear programming algorithms for optimally assigning personnel to jobs. He demonstrated the value of such optimization methods for personnel assignment, despite the fact that the metrics characterizing the test scores were not of the interval type assumed in the derivation of the methods.

Restriction in Range

Carpenter (1970)[2] used simulation techniques to study the distributional sensitivities of the procedures designed to correct statistics for restriction in range. He pointed out that other investigators (e.g., Novick and Thayer, 1969) had studied the adequacy of the restriction in range formulas using real data. When the formulas failed under certain conditions, the researchers hypothesized that the fault was lack of conformity between the statistical assumptions and the population parameters, but were unable to isolate the exact assumption(s).

Carpenter concluded that the correction formulae should be employed by researchers faced with the restriction in range problem unless the departures from the assumptions are extreme.

EVALUATION OF PERSONNEL POLICIES

In large manpower systems (e.g., branches of the Armed Forces), many personnel policies are operative. Possible changes in some of these policies are constantly being considered. Ideas for policy alternatives may

---

[2] Paper presented by G. Rampton at the 12th Annual Conference of the Military Testing Association, September 14-18, 1970.

be externally generated. The concept of an All Volunteer Force and the concern regarding the use of tests for minority group personnel have created a host of policy alternatives for personnel managers to consider.

Conventional Approach

Quite often, a suggested policy alternative has never been tried and, consequently, no historical data are available for predicting the consequences of adopting the new policy. The usual procedure which has been followed in cases of this nature is for the decision-maker to solicit the judgments of experts in the problem area, select a policy alternative and implement it, either on a small experimental group or on a full scale basis. The results of the policy change are monitored and evaluated to provide the decision-maker with feedback on the consequences of the change. In summary, the conventional approach involves two major phases:

(1) Implement new policy; and,

(2) Follow-up and evaluate new policy.

The major shortcoming of this conventional approach is the necessity of implementing a new personnel policy and then evaluating the consequences. Obviously, if the results are favorable, no harm is done and the personnel system benefits. On the other hand, if the consequences are highly undesirable in terms of some criterion (e.g., fleet readiness), the damage will have become an accomplished fact before the subsequent follow-up phase of evaluation informs the policy-maker and the new policy can be rescinded.

Simulation Approach

The undesirable sequence of acting and then receiving feedback on the consequences of the action can be circumvented, with varying degrees of success, using digital computer simulation techniques.[3] Specifically, the simulation approach to personnel policy evaluation would entail four major phases:

(1) Build a mathematical model of the personnel system and simulate various alternative policies on a computer;

(2) Evaluate the simulation results and select the policy which appears promising;

---

[3] This point was made in response to the idea of lowering minority group aptitude test entrance standards for Navy schools, at a recent symposium on minority group testing hosted by the Bureau of Naval Personnel (Sands, 1971b).

(3) Implement the promising policy, either on an experimental or full-scale basis; and,

(4) Follow-up and evaluate the "real-world" results.

The advantages of this simulation approach are considerable. The most obvious benefit to the personnel manager is feedback on the consequences of policy changes prior to implementation.

Less obvious, but quite important, is the ability to examine the measure of system effectiveness under alternative policies applied to the same simulated "persons." This allows differences in the measure of effectiveness to be attributed directly to policy differences.

Finally, the simulation approach permits numerous replications of the same policy alternatives on different computer-generated "persons." These replications can be made rapidly and inexpensively, providing information for:

(a) an assessment of the importance of the chance input of "persons" in affecting the measure of policy effectiveness; and,

(b) an estimate of the variability in system effectiveness which can be anticipated if the new policy is made operational.

In conclusion, it appears that digital computer simulation techniques constitute a powerful tool for the psychologist seeking to aid decision-makers in their difficult task of evaluating alternative personnel policies.

REFERENCES

Alf, E. F. and Wolfe, J. H. Comparison of Classification Strategies by Computer Simulation Methods. San Diego, California: U. S. Naval Personnel Research Activity, Technical Bulletin STB 68-11, June 1968.

Bartholomew, D. J. Stochastic Models for Social Processes. New York: John Wiley and Sons, Inc., 1967.

Boldt, R. F. A Multivariate Function Useful in Personnel Management Models. Washington, D. C.: U. S. Army Personnel Research Office. Proceedings of the U. S. Army Operations Research Symposium, 1962.

Boldt, R. F. A Technique for Simulating Mixed Normal and Uniform Distributions. Washington, D. C.: U. S. Army Personnel Research Office. Research Memorandum 65-9. December, 1965.

Carpenter, G. J. The Effects of Departures from Assumptions on the Restriction of Range Correction Formulas. Toronto, Ontario: Canadian Forces Personnel Applied Research Unit. Paper presented by G. Rampton at the 12th Annual Military Testing Association Meeting, September 14-18, 1970. In: Mahnen, H. A. and Willing, R. C. (Eds.), Proceedings of the 12th Annual Conference, Military Testing Association. Indianapolis, Indiana: U. S. Army Enlisted Evaluation Center, 1970.

Decision Systems Associates, Inc. COMPASS II: United States Navy Computer-Assisted Recruit Assignment Model. Rockville, Maryland: Decision Systems Associates, Inc., September 1969.

Gordon, G. System Simulation. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1969.

Harris, R. N. A Model Sampling Experiment to Evaluate Two Methods of Test Selection. Washington, D. C.: U. S. Army Behavioral Science Research Laboratory, March 1967.

Hatch, R. S. COBRA: United States Marine Corps Computer-Based Recruit Assignment Model. Rockville, Maryland: Decision Systems Associates, Inc., June 1969 (a).

Hatch, R. S. Development of Optimal Allocation Algorithms for Personnel Assignment. Rockville, Maryland: Decision Systems Associates, Inc. Paper presented at the N.A.T.O. Conference held in Portugal. September 1969 (b).

Hatch, R. S. An Ounce of Valid Design Specification is Worth a Pound of Validation. Rockville, Maryland: Decision Systems Associates, Inc. In: Siegel, A. I. (Ed.), Proceedings of a Symposium on Computer Simulation as Related to Manpower and Personnel Planning. Wayne, Pennsylvania: Applied Psychological Services, Science Center, 1971. (In preparation).

Horst, P. A Technique for the Development of a Differential Prediction Battery. Psychological Monographs, No. 380, 1954.

Horst, P. A Technique for the Development of a Multiple Absolute Prediction Battery. Psychological Monographs, No. 390, 1955.

Johnson, C. D. Matching Manpower Resources and Requirements: I. Matching Manpower Resources to Military Requirements, II. Matching Manning Requirements to Manpower Resources. Washington, D. C.: U. S. Army Behavioral Science Research Laboratory. Paper presented at the U. S. Army Human Factors Conference. October 1968.

Johnson, C. D. System Simulation Model Advances for the Evaluation of Alternative Manpower/Personnel Policies. Arlington, Viirginia: Behavior and Systems Research Laboratory. In: Siegel, A. I. (Ed.), Proceedings of a Symposium on Computer Simulation as Related to Manpower and Personnel Planning. Wayne, Pennsylvania: Applied Psychological Services, Science Center, 1971. (In preparation).

Martin, F. F. Computer Modeling and Simulation. New York: John Wiley and Sons, Inc., 1968.

McMullen, R. L. Dynamic Flow Simulation Models. Washington, D. C.: U. S. Army Behavioral Science Research Laboratory. Paper presented at the 23rd Military Operations Research Symposium. June 1969.

McMullen, R. L. SIMPO - I Career-Noncareer Model. Arlington, Virginia: Behavior and Systems Research Laboratory. Technical Research Report 1162. June 1970.

Naylor, T. H., Balintfy, J. L., Burdick, D. S., and Chu, K. Computer Simulation Techniques. New York: John Wiley and Sons, Inc., 1966.

Niehl, E. and Sorenson, R. C. SIMPO-I Entity Model for Determining the Qualitative Impact of Personnel Policies. Washington, D. C.: U. S. Army Behavioral Science Research Laboratory. Technical Research Note 193. January 1968.

Novick, M. R. and Thayer, D. T. An Investigation of the Accuracy of the Pearson Selection Formulas. Princeton, New Jersey; Educational Testing Service. Research Memorandum. RM 69-22, 1969.

Rhodes, K. Background Considerations for Model Evaluation. Washington, D. C.: Naval Personnel Research and Development Laboratory. Special Research Report. November 1970.

Sands, W. A. Cost of Attaining Personnel Requirements (CAPER) Model. Washington, D. C.: Naval Personnel Research and Development Laboratory. Paper presented at the 12th Annual Military Testing Association Meeting, September 14-18, 1970. In: Mahnen, H. A. and Willing, R. C. (Eds.), Proceedings of the 12th Annual Conference, Military Testing Association. Indianapolis, Indiana: U. S. Army Enlisted Evaluation Center, 1970.

Sands, W. A. Determination of an Optimal Recruiting-Selection Strategy to Fill a Specified Quota of Satisfactory Personnel. Washington, D. C.: Naval Personnel Research and Development Laboratory. Research Memorandum WRM 71-34. April 1971 (a).

Sands, W. A. Simulation of the Consequences of Policy Changes. Washington, D. C.: Naval Personnel Research and Development Laboratory. Paper contributed to the Proceedings of a Symposium on Minority Group Testing. Washington, D. C.: Bureau of Naval Personnel. June 1971 (b). (In preparation).

Sands, W. A. Application of the Cost of Attaining Personnel Requirements (CAPER) Model. Washington, D. C.: Naval Personnel Research and Development Laboratory. Technical Bulletin WTB 72-1. August 1971 (c).

Schwartz, M. A. A Ship Simulation Model for Manpower Research. Washington, D. C.: Naval Personnel Research and Development Laboratory. In: Siegel, A. I. (Ed.), Proceedings of a Symposium on Computer Simulation as Related to Manpower and Personnel Planning. Wayne, Pennsylvania: Applied Psychological Services, Science Center, 1971. (In preparation).

Schwartz, M. A., Parker, K. Q., and Rhodes, K. B. Evaluation of a Ship Simulation Model (SHIP II) for Manpower Research. Washington, D. C.: Naval Personnel Research and Development Laboratory. Staff Paper. November 1970.

Sorenson, R. C. Effect of Modification in Operational Procedures on an Optimal Personnel Allocation System. Washington, D. C.: U. S. Army Personnel Research Office. Paper presented at the American Psychological Association Convention, September 1965. Abstract in the American Psychologist, Vol. 20, No. 7, 1965 (a).

Sorenson, R. C. Optimal Allocation of Enlisted Men -- Full Regression Equations vs. Aptitude Area Scores. Washington, D. C.: U. S. Army Personnel Research Office. Technical Research Note 163. November 1965 (b).

Sorenson, R. C. The Effect of Metric Changes on Resource Allocation Decisions. Washington, D. C.: U. S. Army Behavioral Science Research Laboratory. Paper presented at the 5th U. S. Army Operations Research Symposium, March 1966. Proceedings for the United States Army Operations Research Symposium, 1966.

U. S. ARMY CIVILIAN ACQUIRED

SKILLS TESTING PROGRAM

JOHN S. BRAND
U. S. Army Enlisted Evaluation Center
Fort Benjamin Harrison, Indiana

For the past 10 years the Army's Civilian Acquired Skills (CAS) program has been based on the judgments of Classification Interviewers. That is, the possible usefulness of civilian occupational skills of inductees to assignments in related MOS are made by Classification Interviewers during the induction process. In making these determinations, interviewers used the Dictionary of Occupational Titles, job descriptions from AR 611-201, documentation, if any, provided by the inductee, and impressions gained from a relatively short interview. It was normally not possible for Classification Interviewers to acquire any appreciable level of competence in any of the numerous trades and skills which are required by Department of the Army. Experience, therefore, has shown that while existing classification systems have functioned fairly well, these systems were in need of more objective and precise methods of identifying and utilizing previously acquired skills of incoming manpower resources. Too often skills were not being utilized or personnel were assigned to jobs they could not perform effectively. The primary need, therefore, in the Army's induction processes was for superior measurement devices to supplement interviewer judgment. Since the Enlisted Evaluation Center has the mission of personnel measurement and evaluation of enlisted personnel, Deputy Chief of Staff for Personnel, Department of the Army, requested this Center to investigate the feasibility of introducing MOS evaluation tests into the classification process.

Review of MOS tests in the Army MOS structure which are related to civilian occupational skills and trades indicated that selected tests could be used to appraise levels of competence. In the fall of 1970, therefore, the EEC was directed to implement a pilot testing program on 1 January 1971 to introduce the use of MOS evaluation tests as an aid in the Army's classification and assignment procedures.

Thirty-five (35) MOS evaluation tests were selected for the testing of inductees - this phase of the pilot program was designated CAS-I - and 8 tests were selected for use by the U. S. Army Recruiting Command - this phase was designated CAS-JTV (Junior college, technical and vocational schools). Tests included in the pilot program represented 7 of the 10 major areas of the enlisted MOS structure. Typical examples of tests in the program are:

TV Equipment Repairman
Metal Body Repairman
Welder
Carpenter
Mason
Plumber
Electrician
Heavy Vehicle Driver
Computer Systems Operator
Medical Laboratory Specialists
Baker

Tests used in the CAS-JTV program, which were later increased to 11, included some of those in the CAS-I program plus Machinist, Computer Programmer and ADP Systems Analyst.

All tests used in the pilot program were reviewed to eliminate from scoring those items which were specific to the military and would, therefore, not be known to persons from private industry.

Single cut scores were established for both programs to separate personnel into Pass/Fail categories. Cut scores for the CAS-I program were set to approximate the 20th percentile of the Primary MOS distribution, and cut scores for the CAS-JTV program were set at the 40th percentile of the Primary MOS distribution. Cut scores for the CAS-I program were set at a lower level than those for the CAS-JTV program because CAS-I personnel were to be assigned to the MOS in pay grade E-2 while CAS-JTV personnel were to be assigned the entry grade in the MOS - usually E-4.

Test papers were scored at the Reception Station and AFEES, which were provided keys for this purpose. Results were reported by the reception station to Department of the Army by utilizing existing CAS reporting procedures. In these procedures, the classification interviewers indicate estimated level of job skill by the following code:

4. Highly qualified
3. Can be utilized without further training
2. Further training required

For the 35 MOS in the CAS-I program, this code was used to reflect test results, with "4" indicating "Pass" and "2" indicating "Fail". C&I Section interviewers, however, were instructed to continue making estimates of inductee qualifications so that these judgments could be compared with test results. CAS interviews were conducted before testing so that interviewers were not influenced by test results.

Reception stations were instructed to forward test answer sheets and copies of DA Form 20 for each examinee to the USAEEC for program evaluation and analysis. The DA Form 20 includes major background data on each EM to include ACB scores, CAS data to include interviewers rating, education and other pertinent personnel information. CAS test score data was also shown on the DA Form 20.

A follow-up validation program was also built into the CAS pilot program consisting of a special rating form to be completed by supervisors at intervals of 1, 3 and 9 months after unit assignment. Three of these forms were entered into the inductees personnel (201 file) with instructions to Unit Personnel Officers for processing and forwarding to the EEC. The rating form consisted of a simple four-point scale with the following anchors or descriptive statements: Above Average, Average, Below Average, and Unacceptable.

The CAS-I program was initiated 1 January 1971 at the reception station at Fort Knox, Kentucky, and the CAS-JTV program was initiated at 9 Armed Forces Entrance and Examination Stations (AFEES) scattered over the continental US. Due to the initial success and favorable reception of the program at Fort Knox, the CAS-I program was expanded in April 1971 to the remaining 7 reception stations, and 3 additional tests were added to the CAS-JTV program. Strong support for the testing program was indicated by all personnel contacted at the reception stations; support for the program by USAREC appeared to be somewhat less.

The CAS-I pilot study was set up to continue through 30 June 1971 and the CAS-JTV program was set up to extend through 30 September 1971. The testing phase of the CAS-I program was almost wholly terminated by the recent expiration of the draft law. Results from the program are therefore given as of approximately 30 June 1971.

Virtually all of the test results have come from the CAS-I inductee program at the reception stations because of the volume of incoming personnel from selective service. The CAS-JTV program under USAREC has produced negligible test and enlistment results, with 40 possible recruits tested and 8 enlistments. Of the 11 tests in the CAS-JTV program, virtually all of the testing has been in one MOS - 74F, Computer Programmer. The pass rate for this test is about 33% - that is, only 1/3 of the persons tested exceeded the cut score.

Test results in the CAS-I program on the other hand have been very successful. 2800 persons have been tested in 33 of the 35 MOS in the program. The overall pass rate is 60%. By MOS, the pass rate ranges from a low of about 30% in two ADP MOS to a high of 93% for plumbers. 1600 EM have been identified with useable skills. The value of these

skills to the Army in terms of training costs would be around $4,000,000. However, as will be seen, the supply of useable skills provided by the draft exceeds Army requirements, and it has been possible to assign only a fraction of these persons to the MOS in which they were tested.

Comparisons of interviewer codes with test results based on preliminary data thru March 1971 showed a low correlation of .19 (N= 266) which indicates a generally poor agreement between interviewers ratings and test scores. Further analysis of these relationships have not been possible.

Although the pilot program was coordinated at Department of the Army levels, and the special rating forms were placed in 201 files of EM who passed the CAS-I tests, for reasons unknown to this writer these forms were not executed in receiving organizations and returned to this Center as intended. The ratings should have begun coming in to the USAEEC about 4 months after the test date, or during the month of May 1971. The failure of this validation plan to operate and very limited manpower resources at the EEC has curtailed the quantity of follow-up data available for analysis; however, considerable valuable information may be derived from the data that has been obtained. After it became apparent that the ratings were not coming in, efforts were made to obtain assignment information of persons tested under the CAS-I program from both Department of the Army and Fort Knox. Obtaining follow-up ratings was further complicated by the 4-month time interval between testing and rating eligibility and the short time frame for the pilot program. By 30 June the maximum number of EM who could have been rated by unit commanders was only slightly more than 100 cases. Assignment information was obtained for about 85 EM and letters sent to commanders requesting rating data. Responses in percentages were as follows:

Valid ratings obtained	54%
Valid ratings not available (not assigned MOS, AWOL, etc)	15%
No response	31%
	(100%)

Valid ratings were obtained in a total of 17 different MOS. The number of cases per test were too few for separate analysis. The total number of ratings received are summarized as follows in percentages:

Above Average	65%
Average	35%
Below Average	0%
Unacceptable	0%
	(100%)

Thus every individual who passed the adjusted MOS test and was assigned the MOS without AIT was rated average or better by supervisors after 30 days on the job. These results were obtained in spite of relatively low cut scores for the tests as well as the fact that the tests are, in fact, MOS tests and were not designed specifically for the evaluation of civilian acquired skills.

The future of the program depends on the availability to Department of the Army of the rather modest funds required for implementation of a CAS testing program by the USAEEC. It is hoped that these funds may be made available by the last half of FY 72, or by January 1972. An implementation plan has been developed and anticipates the development of 100 CAS tests per year in 4 priority groups. Since the total number of CAS tests will probably not exceed 300, the CAS test program could be largely completed within 3 years. The program will parallel the Enlisted Evaluation System and will utilize most of the EDP and related programs of this system. It is believed that a CAS testing program will increase the objectivity and accuracy of Department of the Army classification and assignment procedures and will also assist in attaining the goals of a Modern Volunteer Army. By scientific utilization of the CAS of its manpower resources, Department of the Army can reduce training costs, optimize personnel utilization, and increase efficiency of operations.

THE CANADIAN FORCES PERSONNEL SELECTION INTERVIEW

Maj. M.A. MARTIN
Canadian Forces
Personnel Applied Research Unit
1107 Avenue Road
Toronto 305, Ontario

THE SELECTION INTERVIEW

"As a Selection device, the interview enjoys unabated popularity". (Wright, 1969). In a 1957 survey by Spriegel and James (cited by Wright, 1969), it was found that 99% of 852 firms interviewed applicants before hiring. An investigation by Shaw in 1968 revealed that 75% of the employers surveyed conducted interviews subsequent to the on-campus interview. However, this and another survey by Rusmore in 1968 revealed that only 7 and 5% (respectively) of the employers had empirical validity to report on the interview as a selection tool (cited in Wright, 1969). Any examination will quickly reveal that very little of the literature provides any quantitative evidence, the remainder dealing largely with opinions, and "how to" topics.

An important departure from the general trend of selection interview literature can be found in the research of Professor Webster and his associates at McGill University. Webster ignored the traditional quantitative topics of reliability and validity in the belief that application of rigorous, experimental examination of the interview would identify the underlying processes which provide the basis for the decision making which goes on in the selection interview. Once these processes were known, Webster felt, "...more accurate decisions may occur if the interviewer can increase his control over the way he arrives at conclusions" (Webster, 1967). Four of Webster's principal findings are:

1. Interviewers develop a stereotype of a good candidate and seek to match interviewees with stereotypes;
2. Biases are established by interviewers early in the interview and tend to be followed by corresponding decisions;
3. Unfavourable information is most influential on interviewers;
4. Interviewers seek data to support or deny hypotheses and, when satisfied, turn their attention elsewhere.

THE CANADIAN FORCES

The selection interview came under close scrutiny in the mid 60's with the unification of the three services, since the development of a single selection and assignment system was required. It was decided then that the interview would be retained for use in conjunction with a general intelligence test for selection, and between 6 and 8 written tests for trade assignment. The interviewers were officers with substantial military experience, a university degree, and both formal and on-job-training in interviewing. Prior to 1969, the interviewer prepared himself for the interview by examining the candidate's documents including academic record, employment record and so on, as well as his scores on the trade assignment tests. The interviewer then conducted an interview which would permit him to evaluate the candidate on a nine point scale for each of four areas:

1. academic achievement;
2. family background;
3. social adjustment;
4. employment background.

Based on these four areas, a fifth and global rating, Military Potential, critical to the accept/reject decision is made. Typically, ratings below "3" on this scale result in non-acceptance.

Unification of the three services was bringing a generally less experienced and more heterogeneous group into the interviewer role. Either improved interview techniques were required or the interview had to be given a less crucial role in the selection process. At the same time, a study yielded some evidence that the interview was making a significant contribution in

the selection of applicants for the university subsidization plans. This and other indirect evidence prompted the decision to attempt to improve the interview procedure.

THE SELECTION INTERVIEW STUDY

The study was conceived of as having at least four phases.

I. The first was the restructuring of the interview itself, particularly along the lines suggested by Webster. The initial step was to "decontaminate" the interviewer by depriving him of applicants' test results and documents until after the interview and ratings had been completed. This change meant that:

a. Biases, particularly negative ones, will no longer be operating from the beginning of the interview and perhaps even earlier. This makes the most of Webster's suggestions cited above by:
   1. forcing the interviewer to utilize interview information in his stereotype comparison;
   2. preventing negative information from tests scores and so on from establishing a premature bias or rejection decision; and,
   3. preventing the interviewer from formulating tenuous hypotheses to be pursued and accepted or rejected, at the expense of other types of information available in the early part of the interview.

b. It would be possible to determine empirically what unique contribution the interviewer can bring to the selection decision independent of test and biographical data.

Interviewers at the 15 Recruiting and Selection units where this experimental procedure was introduced were instructed to conduct the interview in two parts. In the first, the interviewee was told of the interviewer's total lack of information and invited to describe his background in the four areas

to be rated (which were mentioned earlier). Favourable information was thus allowed to emerge early and the difficult-to-reverse unfavourable bias was avoided at first. General prompting only was permitted in this part. In the second part, the interviewer was permitted to question the applicant but encouraged to begin with less important and less threatening questions first and to avoid searching for solely unfavourable information. The experimental procedure was well received by all interviewers.

II. The second phase was to be a reliability study. This phase is now completed and will be discussed shortly.

III. The third phase planned was a validity study. A sample of some 421 recruit applicants interviewed and enrolled under the old interview procedures, and 451 interviewed and enrolled under the experimental procedures was drawn during late 1968 and early 1969. A criterion of survival to training graduation was used and seven disposition categories for non-survivals were established. An obvious problem with this kind of sample is the length of time it takes to follow the sample through training. For this reason, validity data are not available at this time. Other problems in this part of the study include the relative crudeness of the pass/fail criterion and the overlap of the non-survival disposition categories. That is, it is very difficult to know whether the man released as "lacking motivation" lacked it because he didn't like his trade, had family problems, or did not have sufficient aptitude and became frustrated. Such distinctions are crucial if we are to know which failures to hold the interviewer responsible for. Parenthetically, the largest part of our non-survivals come under the category of motivational deficiency with very few classified under aptitude deficiency.

IV. The fourth pahse of this study will be an attempt to develop a stereotype of the ideal recruit as perceived by the interviewers. Development of such a stereotype could lead to a standization among interviewers with

resultant increases in the reliability and validity of the selection interview. Here, as in the validation study, categorization will be a thorny problem since the source of the data will be the interviewers' narratives which accompany the nine point scales.

THE SELECTION INTERVIEW RELIABILITY STUDY

SUBJECTS

A total of 477 applicants were interviewed; 37 officers and 152 recruit applicants were interviewed twice under the old procedures and 67 different officer and 221 recruit applicants were interviewed twice under the experimental procedures. Numbers varied because preselected time periods were specified and only certain selection units participated.

METHOD

Each applicant was assigned randomly to one interviewer who completed an interview and set of ratings. The applicant was then randomly assigned to another interviewer for a second complete interview and set of ratings under the same (traditional, or experimental) procedure. Following the two interviews, one was chosen randomly or "designated" to be documented on the applicants' file.

RESULTS

The data were examined for order of interview and designation effects and none were found; that is, there were no differences in ratings given that could be attributed to a re-interview influence.

It was expected that reliability estimates would be lower for the new procedures. Applicants' documents and test results available to the interviewers under the old procedure only, likely influenced ratings in a consistent, inter-interviewer manner. This was so (See Table I).

Officer Sample - officer applicant interviews tend to concentrate on

educational achievement because most were applicants for university subsidization. This concentration (and the fact that this is a fairly objective area) apparently overcomes any lack of knowledge of documented information and the Academic Achievement rating remained at the same high level of reliability. The Military Potential rating reliability declined slightly. The reliabilities of the other three ratings declined substantially however. These areas, of a more subjective nature than the academic, are evaluated on the basis of information from the applicant in both the traditional and experimental procedures. However, under the new procedures, the interviewer is forced into greater concentration on the more critical academic area.

Recruit Sample - With less emphasis on academic background required for recruit applicants, the Academic Achievement rating reliability dropped significantly as did the Military Potential rating. The three other ratings thus remained virtually unchanged.

At this point the drops in reliability with the experimental technique should not be too alarming since reductions of the reliabilities do not necessarily imply that validities will fall correspondingly, or that validities may not rise, under the new procedure. The question to be addressed now is whether or not this price paid in reliability resulted, in fact, in an increase in validity by permitting the interviewer to make a unique contribution to applicant assessment independent of test and background information.

This completes the description of what has been completed in the selection interview study to date. Many of the questions previously raised are being, or will soon be, attended to in the further phases of this study.

Analysis for phase III, the validity study, is in progress and hopefully will answer the questions about the independent contribution of the interviewer to the selection decision, and about the efficacy of the new interview procedures. Coding and categorization of the narrative descriptions of applicants for the purposes of establishment of a common stereotype (Phase IV) have commenced under the direction of an experienced interviewer. Results of these studies will be made available to interested parties as they become available.

REFERENCES:

Webster, Edward D. Decision Making in the Employment Interview. Industrial Relations Centre, McGill University, Montreal, 1967.

Wright, Orman R. Jr. "Summary of Research on the Selection Interview since 1964." Personnel Psychology, 1969, 22, 391-413.

TABLE I

RELIABILITY ESTIMATES FOR OLD AND NEW INTERVIEW PROCEDURES FOR COMBINED SAMPLES AND FOR OFFICER AND RECRUIT SAMPLES

	COMBINED SAMPLE r	OFFICER SAMPLE r	RECRUIT SAMPLE r
1. OLD INTERVIEW PROCEDURE			
Academic Achievement	.78**	.80	.75**
Family Background	.48	.62**	.43
Social Adjustment	.57*	.75**	.47
Previous Employment	.58*	.58*	.58
Military Potential	.74**	.86*	.69**
2. NEW INTERVIEW PROCEDURE			
Academic Achievement	.65**	.79	.59**
Family Background	.47	.43**	.47
Social Adjustment	.53*	.54**	.51
Previous Employment	.62*	.66*	.59
Military Potential	.56**	.81*	.48**

Significance of the difference between old and new interviews for that sample and rating:

* $p < .05$

** $p < .01$

"CLINICAL EVALUATION AND PREDICTION OF MILITARY EFFECTIVENESS OF NAVAL ENLISTEES"

By

Commander (MC) Alfredo Beyer
Peruvian Navy
Head, Psychological and Psychiatric Screening

Ministerio de Marina
Lima, Peru, S. A.

I N T R O D U C T I O N

Last year the undersigned published a report entitled "The Brief Psychiatric Interview and the Prediction of Military Effectiveness" (1). Data was presented demonstrating experience obtained from satisfactory results following interviews of a group of Peruvian Midshipmen who entered the Naval Academy in 1966. The prediction of this examination as related with their retention after four years of active duty.

The purpose of this paper is to report on the same experience as applied to Peruvian naval enlistees. It should be emphasized that in Peru compulsory military service is determined by Government Regulations in two modalities for draftees and enlistees:

a. In the first case, all youngsters of military age, who have been selected were admitted to the Armed Forces as recruits. After three months of training in the Navy, these selectees go to fleet or naval stations as sailors for twenty-four months, and following this training they may apply to the "Naval Technical Training Center", which is the equivalent to the Enlisted School.

b. In the second case, young civilian men, who wish to pursue a profession as Petty Officer, may also apply at the "Naval Technical Training Center".

As one can readily see these enlistees were volunteers for military service. They qualified according to medical, psychological, moral, and academic standards.

It is necessary to comment that draftees and enlistees have not been subjected to a prior screening as used at Recruiting Stations or Armed Forces Examining Stations.

The enlistees selected must take military and academic training and attend basic schools equivalent to Basic Class "A" Enlisted Schools in the United States Navy (13). In the latter group we have concentrated on the present study.

Concerning the validity of the initial screening interview of navy enlistees in past decades the United States Navy Medical Neuropsychiatric Research Unit, San Diego, California, has pursued an intensive epidemiological investigation of the general effectiveness of naval enlisted personnel (3), (4), (5), (6), (7), (8), (9), (10), (11). In 1963, Plag and Hardacre had established a significant and unique relationship between unsuitability and the predictors of education, age and measure of intelligence (3). The Navy's General Classification Test (G. C. T.), which is routinely administrated to all recruits for classification purposes served in these investigations as a measure of general verbal intelligence. In another report (11), results indicate that, while the initial clinical interview has low but statistically significant predictive validity, its unique significance all but vanishes when it is combined with the variables of age, educational achievement and a measure of intelligence. On the other hand, the interrelationships between clinical prediction and a variety of criteria of military effectiveness are doubtful, as well as the uniqueness and practical value of the initial screening process (4), (8). Plag J. A., Arthur R. J., and Phelan J. D., in recent communication concluded that total four-year rates of military non-effectiveness are relatively constant regardless of whether psychiatric selection is practiced in recruit training or not; and that standard psychiatric selection procedures were minimally valid for lowering attrition and non-effective performance in the fleet; and that a variety of recruit personnel history characteristics were found to be related to fleet effectiveness (7).

For these reasons we recently initiated research on this interesting matter and in the present report we shall attempt to present a preliminary study about the incidence of effectiveness for naval enlistees over four-year separation rates related to initial clinical interview and other variables as age, educational achievement, G. C. T. score, family stability, pathological backgrounds (neurotic and personality traits), and actual symptoms referred to in psychiatric questionnaire, as discussed in the clinical interview by each candidate.

PROCEDURE

Sample: Individuals for this study were enlistees who entered training at Naval Technical Training Center at Callao, Peru, during March 1967 and who remained on active duty for a period of at least four years from their dates of enlistment.

Examination: The examination took place at the Naval Technical Training Center at Callao. All candidates must complete the scheduled following steps:

a. They have to fill out a psychiatric screening questionnaire composed of social history questions and a series of yes-no items referring to psychiatric symptomatology.

b. In addition to several other admission procedures, they must be submitted to a brief psychiatric interview, that is conducted as part of each candidates medical examination. These interviews are conducted by a psychiatrist or by clinical psychologists, while the screening questionnaire served as an aid to the examining clinicians in directing his interview and in arriving at a prediction regarding the candidate's potential adjustment to the naval service. Each subject's predicted service effectiveness is categorized as being: above average, average, below average and marginal or risky.

c. Prior to the interview an intelligence test (GCT), a 100-item USN test of verbal aptitude, was applied experimentally for the first time in Peru, in Spanish translation, with some changes in a few items for cultural comprehensiveness factors. This test was requested officially and its application approved.

Other data such as age, educational achievement, family stability, psycho-pathological traits, are part of the psychiatric screening questionnaire completed by each candidate and explained, as we said above, by each applicant during the clinical interview.

Criterion: Military effectiveness versus non-effectiveness was the dichotomous criterion in this investigation. Sailors were considered to be effective if they completed their four-years of active duty. Those rated as non-effective were enlistees who required early separation from the service. A small number of the experimental subjects were classified as neither effective nor non-effective, having left the

service because of physical disability (excluding all neuro-psychiatric disorders); administrative non derogatory reasons; or death.

Predictors: Predictor variables were obtained from a standard psychiatric questionnaire; from the administrated GCT; and from the data of prediction rating interview. The predictor variables studied were the following:

1. Age at enlistment
2. Years of formal education completed
3. Family stability - the marital status of parents at the time of recruit's enlistment
4. GCT score
5. Initial clinical rating based upon the psychiatric interview, which was dichotomized in two clinical rates: (a) Above average-average, (b) below average-marginal or risky
6. Rates by abnormal personality or neurotic traits that were specified in the psychiatric questionnaire and discussed in the interview.

The statistical analysis: As these comparisons involved discrete data, statistical analysis was made independently for each variable. It consisted of utilizing the Chi-square test for determining the significance of the differences between the predictor variables and the two criteria: discharges (noneffectiveness), versus retentions(effectiveness) in a four-year active duty (12).

R E S U L T S

The total research sample numbered 228 sailors. Of this group 150 rendered effective Naval service, while 65 were classified as being noneffective; 13 were eliminated from the experimental sample as a result of being classified neither effective nor noneffective. These were subjects discharged because of physical disability, excluding all neuropsychiatric disorders; administrative for nonderogatory reasons; or death. These tabulations are shown in Table 1.

Table 1 shows the number and percentage of men who rendered noneffective service and the various parameters of the noneffective criterion.

Figure 1 shows the year of service during which discharge occurred for those noneffective who were separated prior to

the completion of their four-year active obligated duty. It is evident from the data depicted in this Figure, that the largest percentage of noneffective enlistees are identified in the short four week training and the first year of enlisted school.

Table 2 shows that the difference between the two groups about educational level is significant at less than .05 level of confidence. Hence, we conclude that the subjects with higher education level have had more probabilities to retention in military service than subjects with less educational level. In the same Table we depict that there is no significant difference in age at enlistment between the two groups.

Table 3 depicts that the difference between the two groups about initial clinical rates is significant at less than .05 level of confidence. We conclude in this case, that subjects with clinical rates above average and average could have more chance to retention in military service than the subjects qualified as below average and marginal or risky.

Table 4 shows family stability-marital status of parents at the time of the recruit's enlistment. There is no significant difference between the two groups at the .50 level of confidence.

Table 5 depicts about GCT score, dichotomized in 30 or less and 40 or more. There is no significant difference between the two groups at the .50 level of confidence. We also can observe no significant difference of means in GCT scores in both groups (D = 1.30 D - 1.92 z = .66 p = .10).

Table 6 shows the frequencies of 262 positive items of the noneffective group and 616 positive items of the effective group which were categorized as:

a. Actual neurotic traits, i.e.; fingernail biting, hand sweating, etc.

b. Background neurotic traits, i.e.: enuresis more than five years, previous psychiatric consultations, somnambulism, etc.

c. Abnormal personality and character traits, i.e.: easy to anger, running away from home, fits of temper, expulsion from school, etc.

We can observe in the contingency table the expected values for each cell. We conclude that the difference between the number of positive responses in the noneffective with the effective group is not significant statistically.

D I S C U S S I O N

It is interesting to note, despite the small sample, that the incidence of effectiveness for naval enlistees for the entire group of 228 (minus 13 subjects who were neither effective nor noneffective), 28.2 percent were classified as noneffective enlistees, while 71.5 percent rendered effective service, including 5.7 percent classified as having rendered neither effective nor noneffective service. In the last group their service was creditable and for this reason, they could be classified as effective personnel.

Plag, J. A. (11) encountered for a 3,708 man cohort that 72.4 percent rendered effective service while 27.6 percent were classified as non-effective enlistees. We appreciate that there is not much difference in either study. Indeed, various parameters in both incidences are not exactly equivalent, i.e.: we have not been considering the parameter "completed tour but not recommended for re-enlistment" or "unsuitability" among others. On the other hand, excluding physical disabilities, the line Commanders, by means of our regulations, determine all separation for those enlistees via administrative channels for reasons of inaptitude. These are usually disciplinary and academic failures. Hence, Table 1 shows a high incidence of attrition for disciplinary reasons, 47 subjects or 72 percent of all separations. We are sure that if in our regulations were established the Aptitude Board system, many of these could be qualified as "unsuitable". In summary, almost three out of every 10 new enlistees fail to make a satisfactory adaptation to the military environment.

Another finding is the declining rate of attrition through the four years of the first enlistment. It is reasonable to think that the first months of training are a difficult time for adjustment of personality's resources. It is observed in almost all military schools.

It is interesting to note that in the predictors, that were described above, only in the education level and the initial clinical rating, the differences between the discharged and remaining groups are low but statistically siginificant at the .05 level of confidence. The differences with the other predictors in both groups are not significant.

We will try to explain all these findings. The relation between military effectiveness and the educational level is reasonable and it agrees with the results of other United States Naval investigations (3), (4). On the other hand, Plag and Arthur demonstrated that the backgrounds of some neurotic traits or personality attributes have had a poor relation with efficiency during the service (6). Age at the time of enlistment, also had a low correlation among other predictors (3), (4). It is interesting to appreciate that these predictors (family stability, GCT score and initial clinical interview rating), were found to be different from other similar studies (3), (11). This would suggest that the variables indicate a disparity in the sociocultural and environmental aspects. Since family stability was not related, in our study, with effectiveness, it is possible to assume that military life, in countries with a long peace time, could be a factor of protection and security in these subjects. For the GCT score, it is possible that the same cultural factors contribute to the low score, and no significant difference was found between the two groups. In another study, not published yet, we observe that GCT score have had high significance related to midshipmen's academic performance. It is necessary to indicate that all were high school graduates. In the experimental sample a high 52 percent of the subjects have had a lower educational level. At the present time it does no longer occur because applicants must have completed at least the equivalent of 10th grade of high school. The validity of the clinical interview have had low but statistical significance related with effectiveness. The data obtained in midshipmen reconcile these findings. It is necessary to point out that in this procedure, it was considered not only clinical aspects but also a global measurement of personal and sociocultural factors, i.e.: motivation, vocational interest, intellectual potentialities, etc. This global clinical decision, indeed is subjective and for this reason we find technical difficulties, principally in the variability of diagnostic criteria (2), (14). This technical problem could be solved by means of examiners' training, improving methods to make criteria uniform. Its realization is feasible and the results deserve confidence in small number of applicants as occur in our environment. I think that it is necessary to continue this research for improving techniques for personnel assessment.

S U M M A R Y

This study was designed to evaluate the military effectiveness in a group of 229 applicants who entered the Naval Technical Training Center at Callao (Peru) in 1967. For

this experimental sample, it was found that approximately 72 percent rendered effective service. Subjects classified as rendering effective service were those completing four-years of active obligated duty. Military effectiveness versus noneffectiveness was the dichotomous criterion used in this investigation. The comparison was realized upon the base of such predictors as age, educational level, family stability, GCT score, initial clinical interview rating and rates by abnormal personality and neurotic traits referred in psychiatric questionnaire. For statistical analysis data was utilized Chi-square test, that shows in relation to predictor enunciated, only two of them, have had statistical significance at .05 level of confidence. These predictors were initial clinical interview rating and educational level. These results were commented and compared with similar investigations reported.

TABLE 1

INCIDENCE OF EFFECTIVENESS FOR NAVAL ENLISTEES
N=228)

			Number	Percent
1.	Discharge Separation (Noneffective)		65	28.5
	a. Academic failure	9		
	b. Disciplinary (Unfitness, misconduct)	47		
	c. Neurological	4		
	d. Psychiatric	5		
2.	Discharge Separation (Neither effective nor noneffective)		13	5.7
	a. Administrative	7		
	b. Other medical reasons	5		
	c. Death in fleet	1		
3.	Effective Four-year Duty		150	65.8

TABLE 2

FOUR YEAR SEPARATION RATES BY EDUCATION AND AGE

VARIABLE	Total Subjects	Discharged		Retentions		Difference
Education (Level)	N	N	%	N	%	
8th or less	84	34	52	50	33	$X^2 = 6.88$
9th	76	18	28	58	39	df = 2
10th,11th,12th	55	13	20	42	28	p = .05
Sum	215	65	30	150	70	
Age						
16	58	18	28	40	27	$X^2 = 2.18$
17	72	25	38	47	31	df = 2
18+	85	22	34	63	42	p = .30
Sum	215	65	30	150	70	(NS)

TABLE 3

FOUR YEAR SEPARATION RATES BY CLINICAL INTERVIEW

VARIABLE	Total Subjects	Discharged		Retentions		Difference
Clinical rate	N	N	%	N	%	
Above average Average	141	36	55	105	70	$X^2 = 4.25$ df = 1 p = .05
Below average Marginal (risk)	74	29	45	45	30	
Sum	215	65	30	150	70	

TABLE 4

FOUR YEAR SEPARATION RATES BY FAMILY STABILITY

VARIABLE	Total Subjects	Discharged		Retentions		Difference
Family stability (parents)	N	N	%	N	%	
Living together	144	46	71	98	65	
Separated	37	9	14	28	19	X^2 = .8534
Deceased (one or both)	34	10	15	24	16	df =2
Sum	215	65	30	150	70	

TABLE 5

FOUR YEAR SEPARATION RATES BY G.C.T. SCORE

G.C.T. Score	Total Subjects	Discharged		Retentions		Difference
	N	N	%	N	%	
39 or minus	141	45	69	96	64	X^2 = .543
40 or more	74	20	31	54	36	df = 1
Sum	215	65	30	150	70	p = .50 (NS)

DIFFERENCE OF MEANS IN G.C.T. SCORE

G.C.T. Score	Retentions	Discharges	D_M	J_D	z	p
Mean $\pm$ SD	36.8$\pm$11	35.5$\pm$14.3	1.3	1.97	.66	.10

TABLE 6

FOUR YEAR SEPARATION RATES BY ABNORMAL PERSONALITY ATTRIBUTES AND NEUROTIC TRAITS REFERRED IN PSYCHIATRIC QUESTIONNAIRE

Items	Discharges Yes	Retentions Yes	
(a) Actual neurotic traits	(122.6) 126	(288) 285	411
(b) Background neurotic traits	(23.8) 24	(56.1) 56	80
(c) Personality and character attributes	(115) 112	(271) 275	387
	262	616	

$X^2 = .2796$ $df = 2$ $p = .90$ (NS)

REFERENCES

(1) BEYER, A.: The Brief Psychiatric Interview and the Prediction of Military Effectiveness, in "Current Research Techniques in Personnel Assessment", pp 76-83, Proceedings of the 12th annual Military Testing Association, Ed US Army EEC, In 1970.

(2) GLASS, A. J.: et al, Psychiatric Prediction and Military Effectiveness, Part III, Factors Influencing Psychiatrist, US Armed Forces Med. J. 8: 346, 1957.

(3) PLAG, J. A., and HERDACRE, L. E.: The Validity of Age, Education and GCT Score as Predictors of Two Year Attrition Among Naval Enlistees, U. S. Nav. Med. Neuropsychiat. Res. Unit, San Diego, Rep. No. 64-15, 1964.

(4) PLAG, J. A., and HERDACRE, L. E.: Age, Years of Schooling and Intelligence as Predictors of Military Effectivensss for Naval Enlistees, U. S. Nav. Med. Neuropsychiat. Res. Unit, San Diego, Rep. No. 65-19, 1965.

(5) PLAG, J. A., ARTHUR, R. J., and GOFFMAN, J. M.: Dimensions of Psychiatric Illness Among First-Term Enlistees in the United States Navy, Military Medicine 135: 665, 1970.

(6) PLAG, J. A., and ARTHUR, R. J.: Psychiatric Re-examination of Unsuitable Naval Recruits: a Two Year Follow-up. Amer. J. Psychiat. 122: 534, 1965.

(7) PLAG, J. A., ARTHUR, R. J., and PHELAN, J. D.: An Evaluation of Psychiatric Selection at Naval Training Centers. U. S. Nav. Med. Neuropsychiat. Res. Unit, San Diego, Rep. No. 70-30, 1970.

(8) PLAGG, J. A., and GOFFMAN, J. M.: The Prediction of Four-year Military Effectiveness from Characteristics of Naval Recruits. Military Medicine, 131: 729, 1966.

(9) PLAG, J. A.: The Practical Value of a Psychiatric Screening Interview in Predicting Military Ineffectiveness. U. S. Nav. Med. Neuropsychiat. Unit, San Diego, Rep. No. 64-7, 1964.

(10) PLAG, J. A.: Some Considerations of the Value of the Psychiatric Screening Interview. J. Clin. Psychol. 17: 3, 1961.

(11) PLAG, J. A.: A Decade of Research in the Prediction of Naval Enlistee Effectiveness. U. S. Nav. Med. Neuropsychiat. Unit, San Diego, Rep. No. 70-21, 1970.

(12) SIEGEL, S.: Nonparametric Statistics for the Behavioral Sciences, McGraw-Hill Book Co., Inc., 1956.

(13) THOMAS, E. D.: Navy Recruit Classification Test as Predictors of Performance in 87 Class "A" Enlisted Schools. U. S. Nav. Pers. Res. Activity, San Diego, Res. Rep. SRR 69-14, 1969.

(14) WALLINGA, J. W.: Validity of Psychiatric Diagnosis. U. S. Armed Forces Med. J., 7: 1305, 1956.

SYSTEMS APPROACH TO EVALUATION AND QUALITY CONTROL OF TRAINING

LTC BRYCE R. KRAMER
Chief, Evaluation Division, Office of the Director of Instruction

and

RICHARD S. KNEISEL
Special Assistant - Educational Advisor
United States Army Infantry School, Fort Benning, Georgia 31905

Paper Presented by LTC KRAMER

BRIEF

(SLIDE 1) *Ladies and Gentlemen, as indicated by the title of this presentation, my primary concern this morning is to describe a system for Army Service School course design, a formalized model for implementation of this system which is in use at the US Army Infantry School, the application of this model to a specific course of instruction, the outcome of the application of a systems approach to evaluation and quality control, and the implications for training per se.

INTRODUCTION

To paraphrase William Shakespeare (SLIDE 2) "All the world's a system and we the men and women are merely analysts." Theoretically, the world may be considered an orderly overall process with we, it's inhabitants, attempting to understand the process and, in some measure, do something about it. Within the overall, larger process are smaller processes or systems and within these smaller systems are still other ones, wheels within wheels--or perhaps (SLIDE 3) shinbones connected to leg bones, leg bones connected to thigh bones and all the bones ultimately connected to the head bone. So it is with training--we hope, not merely a skeleton but a system. Somewhere within this worldly process there is a milieu called, loosely, education and training, within which is a smaller cosmos of military training. Within military training is Army training (US-type, of course). Within the Army training is that which we call Infantry and lo! within the Infantry training is the Infantry School (SLIDE 4). Some persons might conclude that this represents the foot bone but we of the "Follow Me" School prefer to believe that we are more logically the head bone of the system.

*Slides are found beginning on page 133

SYSTEMS ENGINEERING OF TRAINING (COURSE DESIGN)

As we approach training within a Service School, such as the Infantry School; in order to be able to attack the total problem, it is necessary that a division of labor or some smaller organizational structure be undertaken. This generally falls into place by addressing the total training problem in the development of specific courses of instruction which are designed to provide for qualified graduates to meet stated needs of the military establishment. To insure that these needs are, in fact, being met there is a further requirement to establish some sort of standard and an evaluation or quality control system for this macrocosm of the military training world. The systems engineering process in use in the US Army Service School system is the terminology for this process. This approach represents the logical development of the instructional program that is to meet the Army's needs in a given element of the training system. Other establishments, be they military or civilian, perhaps have other semantical titles for the logical process and systematic approach to the development of the training programs. Be that as it may, an explanation of the US Army Service School systems engineering process, and the logical steps involved, appears to be in order (SLIDE 5). It consists of these seven elements:

1. Job Analysis
2. Selecting Tasks for School Training
3. Training Analysis
4. Preparation for Training
5. Developing Testing Materials
6. Conduct of Training
7. Quality Control--Feedback

A quick look into each of these will help to understand the system and the interrelationship of the elements and, at the same time, understand how this approach lends itself to interrelating with the overall evaluation and quality control of training.

Job Analysis

The Army trains the soldier to perform in a Military Occupational Specialty (MOS) and these MOS's may contain one or more different jobs. The first step in systems engineering is to perform a job analysis. This identifies the on-the-job performance requirements in terms of individual task and job characteristics of the MOS; such as, duty, position, work environment, and equipment requirements. The completed

job analysis sets the framework within which all subsequent steps of the systems engineering process occur. This basic framework is task-based and job-oriented. Emphasis is on identifying the specific job requirements--those observable acts and behaviors required of MOS incumbents. This step consists of identifying the job and developing the task inventory.

Selecting Tasks for Training

The second step in the systems engineering process is to select from the task inventory a list of those tasks that require, or should receive, formal school training. This implies other selections as well. For instance, tasks not selected for school training obviously are selected for training elsewhere or are specified as prerequisites. (Essentially, the prospective student has three main opportunities to learn to perform tasks: he has already learned to perform certain tasks; he may receive school training on the tasks in a course of training; or, he may learn the tasks on the job.) Which tasks he learns under which conditions must be ascertained. The major consideration for the Service School in this process is to identify those tasks most essential for formal school training. Special considerations--all of which are judgmental--such as the following, must be evaluated and decided upon in varying degrees to come up with the ultimate decision. These include task criticality, task similarity, tasks essential for other tasks, prerequisite ability, capability of learning tasks OJT, time available to develop competence, and percentage of persons performing the task.

Training Analysis

Training analysis is the third step in the engineering of a course of instruction. It bridges the gap between the job requirements and the classroom. The indication of the job in task statement form and the selection of tasks for school training up to this point represent a level of generality that is too gross for instructional purposes. Hence, further analysis of each task selected for training is needed before actual preparation of training materials and test instruments can be undertaken. This procedure of bridging the gap involves:

a. Identifying the job conditions, standards, and supporting skills, knowledges, and attitudes.

b. Converting the job requirements to training objectives and criteria.

c. Developing course structure.

d. Developing course evaluation concept.

Preparation for Training and Testing

Following the training analysis we move into steps 4 (Preparation for Training) and 5 (Testing) of the system. These steps, while indicated independently in the formal breakout, really go on at the same time and represent an interchange between these two elements. Let us discuss first the preparation for training which we may term the production phase of the systems engineering process, because all instructional and administrative materials are developed during this phase to include the course of instruction which we refer to in the Army as the Program of Instruction (POI). Also developed at this point are the lesson plans, handouts, and the training media to support the learning. Regardless of the kind of instructional materials being produced, preparation is governed by student-centered learning principles, which guide him toward the successful accomplishment of the training objectives. This step includes the listing of teaching points, references, methods of instruction (conference, case study, practical exercise, etc.), media (TV, programed text, simulators, models, etc.), the training equipment material (ammunition, trucks, generators, etc.), the facility (field, laboratory, mobile repair shop, etc.). These actions--integrated--lead to the POI and the Training Schedule. As I indicated earlier, the Testing phase is being developed at the same time as the Preparation for Training. In this phase we are concerned with the measurement of the achievement of the individual. That is, we want to know what the student is to learn in the course and evaluate student accomplishment in terms of the training objectives as specified in the criteria set forth in the earlier training analysis. We are not concerned with ranking and grading of students, but rather with a score to evaluate whether the student can or cannot do the job as we set it up in the individual objectives. All the principles of testing and evaluation hold in this step--plus another one or two--mostly a change in perspective to the fact that the test is an instrument to evaluate the program and not the student.

Conduct of Training

Up to this point in the system we have been concerned with getting ready to present the instruction. All of the preceding actions involve the myriad of decisions with regard to what should go into the curriculum. All of this by way of getting the course ready to be taught--this is the

sixth step--that is, the actual instruction or the conduct of training. This aspect, while in a sense independent of the curriculum, does fit in with the accomplishment of the mission, for it takes all of the preparation and translates it into the action element. It involves the instructor's direct action in getting the objectives from the here to there--from the planning to the doing. In this phase the mechanics of the presentation, the facilities, training of the instructor, and general on-going environment of the training are concerns. So, we have finally taught. Are we through--no, we go to the seventh step.

Quality Control of Training

The seventh step in the process of systems engineering is the trial and evaluation, or quality control, of the instructional system. This aspect of the system must be viewed as a continual, empirically-based process which consists of analyzing various feedback data and adjusting the instructional system to insure that the basic objectives of the course are being met. It involves the injection into the system, from time to time, of new data--new developments, for example, that are brought on by changes in technology and doctrine. Quality control has one basic objective of insuring a predetermined quality of training produced through an instructional system that represents the optimum and most efficient mix of instructional resources. Even the most thorough and professional systems planning and design represent only the best prediction or estimate of what will happen when the instructional system is implemented. The training manager starts with the assumption that systems adjustments, even major ones, will be necessary. The instructional system is developed through job analysis, selection of tasks for training, specification of performance-stated training objectives, development of training materials; and the development of measures of proficiency. Quality control continually examines and adjusts, as needed, all elements of this system so as to produce the desired quality of training with the least possible expenditure of resources. This quality control represents a continual feedback through internal sources; such as, tests, evaluation of instructors, opinions of students and through external sources; such as, surveys, research, field review, changes in doctrine and new equipment. This continual feedback affects any one or all of the elements of the system.

USAIS MODEL (SYSTEM) FOR QUALITY CONTROL OF TRAINING

As indicated in the discussion on the systems engineering process, and in conjunction with the foot bone, head bone concept, there is a smaller element of internalizing and formalizing the quality control

and evaluative process that addresses the instructional system and the individual instructional program. A formalized model (SLIDE 6) eliminates many of the difficulties and the confusion that generates in a large educational institution. The model that appears to be most satisfactory for the Infantry School, and hopefully for other Service Schools, involves the concept of establishing the framework and organization necessary for implementing the system and providing for the proper feedback within the system.

The United States Army Infantry School model for evaluation and quality control in the minds of many could best be the one of the young lady you just viewed. Nevertheless, in the hard world of the practical day-to-day fighting of the quality control battle, the model looks more like this diagram (SLIDE 7).

The United States Army Infantry School is, as most organizations and certainly most military Service Schools, one that is essentially a line and staff configuration. The line elements are the instructional departments or divisions, such as the Leadership Department, Brigade and Battalion Operations Department and the Airborne Department; and, the staff being, for example, the Director of Instruction, Director of Operations and Logistics and Chief of the Office of Management and Budget. The quality control model depicted here is being used with the existing organization at the Infantry School. Moreover, we believe it can be used with any existing educational institution.

Perhaps, as you look at this system and as we go through its elements, you may be struck with the idea that this represents nothing new. If I listen carefully I can hear such comments from the audience to the effect that "We do this all the time in our training establishment." Be that as it may, the US Army Leadership Factory--by its official title, United States Army Infantry School--felt the same way for quite a long time. However, there were many pieces of the instructional system which were not meshing and there were quite a few aspects which were "falling through the cracks." Consequently, it was decided to look at the total process, formalize it and integrate it (perhaps impose it on the organization are more appropriate words) within the existing organizational structure. Hence, the Infantry School established this system or model, if you will, for the quality control of the United States Army Infantry School instructional process. This model is one which we feel can have rather universal application.

Before I go into a particular application of this model to an Infantry School course of instruction--namely, the Infantry Officers' Basic Course (IOBC), I believe it would be in order to give a few words of explanation about each of the elements. It would be well to say here that even though there are specific labels of the parts of the system, these parts are not necessarily purely discrete. As in any process, the separate elements tend to flow into one another and in many instances are cross related. Perhaps if it were possible, the model might take on a sort of three dimensional quality. However, in order for us to look at the appropriate "shinbones" and "leg bones," we have made our paradigm two dimensional so that we could discuss it without too greatly complicating the concept. Therefore, I will discuss briefly what is encompassed by the elements of the system--analyze it if you will.

Concept

Flowing out of the overall process of the systems engineering of instruction described earlier is the CONCEPT of how to evaluate the instruction--the course itself, the instructional system implementing the course, and the student undergoing the training. The CONCEPT develops logically in the actual designing of the course, but it becomes structured when the group in the training institution--in the Infantry School it is the Curriculum Planning Committee (appropriate department heads and staff elements)--sit down and specifically indicate just what the course is to do, just what the quality control aspects will be, and how they will take place. For instance: Will there be a diagnostic testing program? Will there be an overall comprehensive test? Will there be end-of-block examinations? Will there be a peer evaluation? What are the elements needed to insure that the system is functioning and to insure that the graduates are, in fact, able to perform? How will we evaluate the training objectives of the course designers? What sort of student questionnaires are needed? What sort of follow-up to the field is desired? What methods for assessing student confidence can be used? How much computerization of the evaluation instruments will be needed? These "for instances" are but some of the aspects that are generated in the CONCEPT phase.

Plan

In conjunction with and logically flowing out of the CONCEPT, comes the PLAN. The establishment of the specific measures in a

concrete form and the time-phasing of these measures (which may involve some PERTing) constitute the PLAN. Perhaps some organizations may include the PLAN in the area we have labeled CONCEPT. However, the PLAN, as envisioned here, is the sum total documentation of the actual finite actions that derive from the conceptual ideas. It is the actual framework of the quality control system. It involves the specifics of who, what, where, when, and how. It specifies the responsible agencies for each of the parts of the quality control effort; it designates the managerial and execution aspects of each of the actions; and, it puts substance on (makes concrete) the conceptual overview. The PLAN, in fact, is a formal document that is issued in the name of the Assistant Commandant.

Execute

The directive (PLAN) having been finalized and issued must then be implemented. Each of the action agencies place into motion their portion of the PLAN. Each of the agencies follows through with what it is supposed to do. For example, if a military stakes performance examination is to be the end-of-course mechanism, the PLAN will have designated an instructional department as the action arm for putting together the examination and making all the administrative and logistical arrangements necessary for conducting the test. The execution involves internal systems for specific aspects of the PLAN.

Monitor

Although this element of the quality control model is indicated here as following the EXECUTE element, it is in a sense a pervading one that touches all the other elements of the model. Most of the MONITORING action comes with the EXECUTION phase and in the follow through of the subsequent activities. MONITORING carries with it the requirement for a specific managerial agency. At the Infantry School, the Office of the Director of Instruction is the project manager for the Quality Control System. This staff office is charged with insuring that the particular actions are accomplished at the proper time and that all aspects of the plan are being integrated and followed through. The project manager keeps abreast of the EXECUTION to insure that appropriate progress is being achieved and he incorporates any modifications that may be necessary. He insures that there is no degradation of the overall concepts of the quality control framework and insures a logical, orderly, and timely flow of data.

Collect Data

This flow of data may be termed COLLECTION of DATA. DATA COLLECTION may appear to be fairly mechanical. Whereas this is generally the case, it is this aspect of the system that many times breaks down just because of that fact. All of the earlier elements of the model can go for naught if this rather mundane aspect is not meticuously attended to. It involves the complexities of manual processes and computer activities. It requires a system of academic record keeping for the student as well as for the system. The data must be in a form useable throughout the instructional institution. The DATA COLLECTION has both immediate and long term implications. It is the basis for the ANALYSIS.

Analyze

Once the data are COLLECTED, they must be manipulated, massaged and interpreted by the project manager. Appropriate statistical measures, such as t-test and chi-square correlations are made and the data reduced to manageable and understandable format. The ANALYSIS of the data ascertains if, in fact, standards are being met. It reveals weaknesses and strengths of the course and the instructional system, and puts the quality control efforts into a meaningful form for the decision-makers to use. It results in a report that provides for a means of appropriate FEEDBACK into the quality control process, and into the larger process of systems engineering of the course and still larger aspects of the total instructional establishment. ANALYSIS provides the mechanics for the FEEDBACK.

Feedback

The ANALYZED data is translated to the students, the managers, operators, decision-makers, planners, and the total system. This is FEEDBACK. It is internal to the quality control, and external to the course and the larger system and the institution, as well as to higher headquarters. The FEEDBACK is the process that forms the basis of upgrading all aspects of the institution, and for the redesign of the course and all the interrelating systems.

Redesign

Perhaps to some, the element that has been labeled REDESIGN is, in fact, FEEDBACK. The intent here is to indicate that something is actively being done with the FEEDBACK. It points to the dynamic and viable quality of the model. Although there is an indication by the term REDESIGN that there is a wholesale change, this is not the

intent. The REDESIGN addresses those aspects that, in fact, warrant some change. In some instances, the analyzed feedback data is confirmatory and in others it points to a redirection and a restructuring, both internally and externally.

A CASE IN POINT - INFANTRY SCHOOL QUALITY CONTROL MODEL APPLIED TO THE INFANTRY OFFICERS' BASIC COURSE

To help you understand this quality control model in operation, I shall run briefly through what transpires in the systems approach to evaluation and quality control of training as applied to one of the United States Army Infantry School Courses--namely, the Infantry Officers' Basic Course (IOBC) (SLIDE 8).

Concept

Training newly commissioned officers has historically been one of the most nagging problems that has faced the Infantry School. It had gone unsolved for nearly 19 years due primarily to the fact that attempts to solve the problem had addressed the adjustment of the length and content of the course rather than restructuring the course as a whole. We recognized that the existing course, while adequate, was certainly far from the outstanding course that we wanted. In March 1970, the Infantry School set aside other pressing projects and focused the attention of its most talented people on the Infantry Officer Basic Course. The Director of Instruction and Department Directors, assisted by many younger officers, who had recently returned from the field, spent 5 continuous weeks in developing a new course under the guidance of the Commandant.

During the CONCEPT development (SLIDE 9) all preconceived notions about basic officer education and training were discarded and this new course was systems engineered from the ground up. New educational and training concepts were developed, and a course was conceived that encompassed the following: diagnostic testing, performance-oriented training, phased instruction that ran the gauntlet from the introductory to circular proficiency courses (CPC's) which simulated a day in combat, specialized training, performance testing, peer ratings, confidence, quality of instruction and attitude questionnaires, a tailored testing and evaluation program that addressed all aspects of the learning situation; and the conduct of a controlled experiment to test the validity of the new course.

Plan

Upon completion of the systems engineering process and CONCEPT development, the Director of Instruction was tasked with finalizing

the PLAN for the 12-week, 4-phase course (SLIDE 10). Phase I, the Introductory Phase, was a 7-week block which included in-processing, classroom, physical training, and range work. During Phase I, all of the theory and other fundamental instruction was programed for presentation. It involved a total of 4 weeks of indoor and 3 weeks of outdoor instruction. Within the classroom portion of the instruction, emphasis was to be placed on practical application situations. A minimum of 1 hour each day was devoted to physical conditioning type training to prepare the student for subsequent phases. Phase I terminated with a comprehensive examination.

Phase II was 3 weeks in duration and was devoted solely to practical application. During this phase, the students were to be organized into platoons and run through tactical training exercises which we called "continuous performance courses" (CPC's). These courses were from 3 to 10 miles in length and included 5 to 10 tactical situations, on each course. Leadership positions were to be rotated upon completion of each requirement. In addition to tactical training, maintenance, medical service, communications, fire support, airmobile operations, mechanized Infantry, and many other combat support activities were integrated in the CPC's. During this phase, over 3,000 leadership opportunities were provided for each class. This phase was a learning as opposed to an evaluation phase and stressed practical application instead of theory. Student peer ratings were integrated as a motivational factor with no other grade being assigned during this phase. School cadre were designated to control the training and conduct short critiques at the conclusion of each exercise.

Phase III was 1 week in duration and devoted to Ranger type training. After an introductory period, the students would execute a 96-hour continuous Ranger problem which emphasized patrolling, raids and ambushes. As in Phase II, a student peer system was used for a motivational mechanism.

Phase IV, the Evaluation Phase, was 1 week in duration. It consisted of a professional facts examination, written performance situations from Phases II and III, land navigation, a comprehensive examination; and a 2-day Military Stakes Examination that consisted of 47 performance situations, such as those portrayed by this film (30 seconds of Military Stakes Film).

At the conclusion of the PLANNING Phase, the Director of Instruction published 2 documents that served as the basis for conducting the experimental course. The first document addressed the content,

sequence, and responsibility for teaching each training objective in the 4-phase course. It addressed the specifics to be taught, responsible departmental agency and the requirement for the formulation of lesson plans, vault files, support requirements, etc.

The second document addressed the overall quality control and the managerial aspects of execution. It encompassed such things as the overall project managership of the test and evaluation program, review of examinations and performance tests prior to their administration to students, training inspections, and proponent agencies for execution and monitoring of the controlled test.

Execute

Upon receipt of the PLAN, the Infantry School staff agencies and instructional departments were given 60 days to prepare for the conduct of the new course (SLIDE 11). The implementation and conduct of the controlled experiment encompassed a 6-month timeframe during which 2 control and 2 experimental IOBC classes, taught under the new concept, were conducted. During the EXECUTE phase, the proponent agencies for the conduct of the instruction and the testing and evaluation of the programs executed their detailed responsibilities. The 7 instructional departments conducted their respective portions of the Phase I training, which terminated with the Phase I comprehensive examination administered by the Brigade and Battalion Operations Department. The Company Operations Department, proponent for the conduct of Phase II training, controlled the overall administration of the CPC's used as training vehicles in Phase II, and incorporated other departments into this training through the use of the visiting department concept. The Ranger Department had proponency for and conducted the Phase III training. The Leadership Department designed, administered, and processed student peer ratings which were used as a motivational factor in Phases II and III.

In Phase IV, the test and evaluation phase, proponent departments administered the various examinations in this phase, i.e., the Company Operations Department, which had proponency for the Military Stakes Examination, coordinated the inputs from other departments and conducted the 47 station performance examination. Concurrently, the school staff offices implemented the various managerial actions designated in the test and evaluation program, and provided the resources necessary for the conduct of the training. The EXECUTION Phase brought into play the internal school systems necessary to accomplish this specific aspect of the PLAN. This effort required the coordinated actions of all agencies in the Infantry School.

Monitor

Although listed sequentially after EXECUTION, the MONITORING aspect addresses all elements of the model once the PLAN is published. In the case of the Experimental Infantry Officers Basic Course (EIOBC), the Director of Instruction was tasked as the project manager (SLIDE 12). The function of the project manager was to insure that the CONCEPT, as documented in the PLAN, was in fact translated into reality in the conduct of training and that the results achieved were in consonance with the training objectives initially specified.

The action agencies of the Director of Instruction came into play in the following areas. The Curriculum Division reviewed the finalized POI's while the Evaluation Division reviewed the examinations for the new course prior to their being administered to the students. The Instructional Methods Division conducted periodic classroom inspections during the EXECUTION Phase, while the Curriculum Division course monitor physically observed training. Concurrently, the Evaluation Division reviewed examination results, faculty observations, and student critiques from the test and control classes. In the MONITORING Phase, deficiencies were anticipated, identified and corrections made as required. This aspect is critical in that the project manager must continually keep abreast of the actions being executed by the respective instructional departments to insure that there is no overall degradation between the concept of training and the real world applications.

Collect Data

The PLAN for EIOBC designated the Evaluation Division of the Office of the Director of Instruction as project manager for the DATA COLLECTION effort relevant to the conduct and evaluation of this course. (SLIDE 13) The responsibilities of the project manager encompassed coordination with all agencies involved in the conduct of the course. These responsibilities also covered the procurement of the results of all examinations, questionnaires, performance ratings and other instruments specified in the test and evaluation plan.

During the conduct of the 2 control and 2 experimental classes, the Evaluation Division compiled the data, subjected it to predetermined ADP applications and maintained statistics in a useable form that provided the basis for the ANALYSIS. The performance of this onerous task was critical in that minute attention to detail in the selection and manipulation of data was necessary so as to provide a valid data base upon which to make value judgments during the ANALYSIS Phase.

Analyze

The ANALYSIS of the data COLLECTED during the conduct of the EIOBC evaluation encompassed a correlation of the results of student performance and responses on the multiple evaluation instruments inherent in the test and evaluation program for the course (SLIDE 14). The Evaluation Division, Office of the Director of Instruction, was the project manager for this aspect of the test.

The COLLECTED data was addressed from 2 points of view. Initially, a correlation was made of the results of aptitude, precourse tests, subjective, objective, and performance examinations, faculty inspections and observations, and the student responses to the various questionnaires and peer ratings for the respective courses. This was done to determine the strengths and weaknesses of the courses, by subject area, and to make judgments as to whether the specific training objectives in the 2 courses were, in fact, being met.

The second phase of the ANALYSIS addressed a comparison of the data derived from the execution of the respective courses. The intent was to make a comparison to determine which course produced the most qualified officer graduate. This comparison produced interesting results. The control classes evidenced a significantly higher entry level of prior academic achievement and military knowledge, which was due primarily to their class profile which included a higher percentage of ROTC DMG's. However, the experimental classes scored higher overall averages on the objective, subjective, and performance examinations administered to both classes. The experimental classes also had a significantly higher level of confidence in their ability to perform selected key tasks taught during the course, and rated the overall quality of instruction higher than did the control classes.

The ANALYSIS of the data procured as part of the test and evaluation program for the conduct of this experiment enabled the Infantry School to make an unqualified determination that the experimental course was a more viable and valuable course in preparing officer student graduates to function as platoon leaders; that it was more challenging, conducted at a better pace, and less repetitive of pre-commissioning training than the regular Infantry Officer Basic Course.

The systematic COLLECTION and ANALYSIS of data, in consonance with the overall test and evaluation program, enabled the Infantry School to conclude that the differences in the performance of the students who attended the regular and experimental IOBC's could be attributed to certain basically inherent differences. These differences lay in the respective course POI's, methodology, and instructional tracks followed.

The ANALYSIS was concluded by the publication of a comprehensive report that addressed all aspects of the conduct of the test program and provided the basis for the FEEDBACK.

Feedback

Inherent in the systems approach to evaluation and quality control is FEEDBACK (SLIDE 15). The Evaluation Division of the Office of the Director of Instruction was designated as the project manager to insure dissemination of the FEEDBACK from the IOBC test (a comprehensive, statistical report and summary) to the ASSISTANT COMMANDANT, staff offices, instructional departments, and students. The FEEDBACK provided a basis for the institution to assess the strengths and weaknesses of the courses, as well as the successes and failures in meeting training objectives. It also provided one comprehensive document that addressed all aspects of the experiment which served as the basis for upgrading the courses of instruction and subsequent REDESIGN. Additionally, the published report compiled during the FEEDBACK phase, served as a substantiating document for the Infantry School's recommendation to Continental Army Command and Department of the Army that future Infantry Officer Basic Courses be conducted in accordance with the revised Infantry Officer Basic Course POI.

Redesign

The REDESIGN phase is the pay-off of the systems approach to quality control in service school courses of instruction. All aspects of training are geared toward capitalizing on the experiences gained. The data obtained during the conduct of the Infantry Officer Basic Course experiment was disseminated to all school agencies and time allowed for them to conduct their respective analyses. Subsequently, the Director of Instruction (SLIDE 16), the project manager for REDESIGN, reconvened the curriculum planning committee and REDESIGNED the course as required. The REDESIGN addressed such areas as reducing POI time allocated to areas found to be repetitive of pre-commissioning training, allocation of additional POI time and resources to areas where training objectives were not substantiated by student performance and the integration of educational innovations and techniques to enhance learning.

The outcome of the REDESIGN was a revised 12-week IOBC that we were convinced had solved one of our toughest curriculum planning problems, that of how to branch qualify a newly commissioned officer without "turning him off" in the onset. We were convinced that approval and implementation of this new course would start the newly commissioned officer on his Army career with a challenging and stimulating experience.

The new Infantry Officer Basic Course required additional personnel and resources to support the performance-oriented nature of the course. The decision to implement was deferred by Department of the Army for 1 year as a consequence of budgetary and personnel limitations.

Consequently, the curriculum planning committee reconvened and used the FEEDBACK obtained during the conduct of the experiment to design a 9-week course that capitalized on the strengths of the 12-week course and resulted in a vastly improved course of instruction which is currently being presented to newly commissioned officers.

The experience that the Infantry School gained while conducting the test and evaluation of the new IOBC validated the thesis that a quality control model is a necessary ingredient in the management of Service School courses of instruction.

I trust that by following the special application of the quality control model for the Infantry Officer Basic Course, you have been able to understand what happened at the Infantry School. There may appear to be distinct entities when, in fact, this model represents a process - a viable following and integration of actions (SLIDE 17). The Infantry School quality control model is a totality that, while imposed on an existing organization, does work. Quality control as practiced at the Infantry School is not a distinct and separate aspect of the instructional system, but rather, one that is all-pervading. It provides for the control, modification, and upgrading of the instructional program, and further provides a means for the training manager to engage in sound decision-making.

IMPLICATIONS OF APPLYING THE QUALITY CONTROL MODEL

What are the implications of applying a Systems Approach to Evaluation and Quality Control of Training--the United States Army Infantry School's Quality Control Model? Perhaps no more than that a logical system, a formalized and structured process, has been and can be imposed on an existing organization which is engaged in training. The key to the concept is that it is systematized. It represents a logical flow from the course design. It provides for a specific plan and a project managing agency to hold it together, analyze the data, and return the feedback into the system (SLIDE 18).

For the United States Army Infantry School it has meant better courses, a truer evaluation of the student, and a means for restructuring the total School. It has helped in such projects as the Volunteer Army (VOLAR) and the efforts to individualize instruction, such as Self-Pacing Instructional Text (SPIT) and the Individualized

Learning Center (ILC). The Quality Control Model has tightened up all aspects of the evaluation process. It has aided in making the instruction more student-centered and moved it away from the instructor-centered. Above all, it has eliminated the fly-by-the-seat-of-the-pants system that in fact was no system.

IMPLICATIONS FOR TRAINING

(SLIDE 19) What are the implications beyond the United States Army Infantry School? We leave that up to you who make up a part of the System of the World--the Instructional Macrocosm. Hopefully our experience and our model may in fact have some utility for your Instructional Institutions.

REFERENCES

Annex Q, Army Schools Curriculum: Administration and Training Policies, CON Regulation 350-1, Training: CONARC Training Directive, Fort Monroe, Virginia: Headquarters, United States Continental Army Command, 1969.

Briggs, Leslie J., Handbook of Procedures for the Design of Instruction. Pittsburgh, Pennsylvania: American Institutes for Research, 1970.

CON Regulation 350-100-1, Training: Systems Engineering of Training (Course Design). Fort Monroe, Virginia: Headquarters, United States Continental Army Command, 1968.

Final Report of USAIS Experimental Infantry Officer Basic Course Evaluation. Fort Benning, Georgia: Headquarters, United States Army Infantry School, 1971.

Popham, W. James, & Baker, Eva L. Establishing Instructional Goals. New Jersey: Prentice-Hall, Inc., 1970.

Popham, W. James, & Baker, Eva L. Planning an Instructional Sequence. New Jersey: Prentice-Hall, Inc., 1970.

Popham, W. James, & Baker, Eva L. Systematic Instruction. New Jersey: Prentice-Hall, Inc., 1970.

USAIS Regulation 350-100, Education and Training: Systems Engineering of Training (Course Design). Fort Benning, Georgia: Headquarters, United States Army Infantry School, 1969.

SLIDES

SLIDE 1 - Title Slide

SLIDE 2 - William Shakespeare

SLIDE 3 - Skeleton

SLIDE 4 - The Infantryman Statue

SLIDE 5 - Systems Engineering Process

SLIDE 6 - The Model

SLIDE 7 - Quality Control Model

SLIDE 8 - Transitional Slide-Infantry Officer Basic

SLIDE 9 - Concept Development

SLIDE 10 - Planning Objectives

SLIDE 11 - Execution Requirements

SLIDE 12 - Monitoring Requirements

SLIDE 13 - Data Collection Requirements

SLIDE 14 - Analysis Requirements

SLIDE 15 - Feedback Requirements

SLIDE 16 - Redesign Requirements

SLIDE 17 - Quality Control Model

SLIDE 18 - USAIS Applications of Quality Control

SLIDE 19 - World-wide Applications

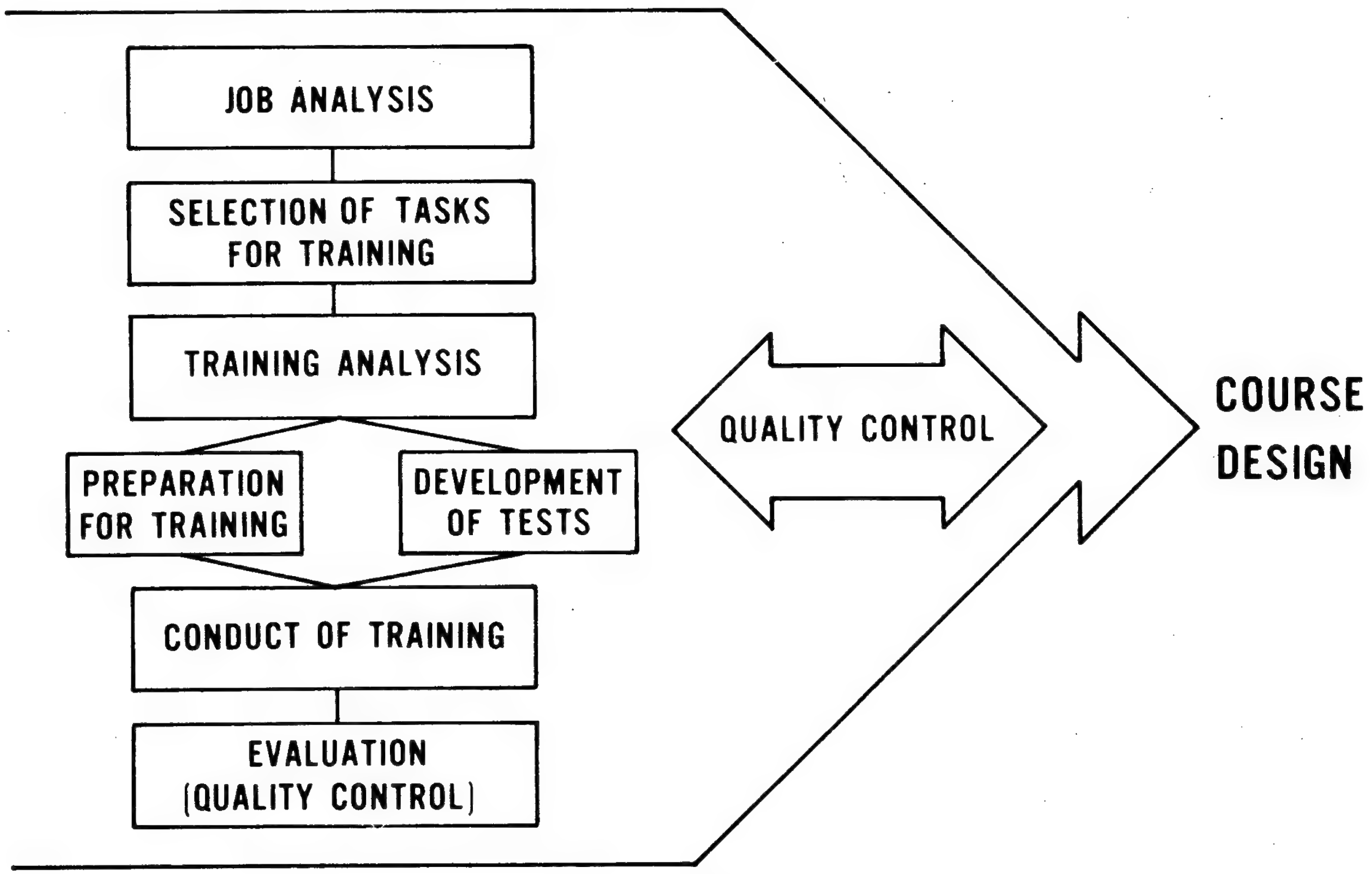
SLIDE #1

SYSTEMS APPROACH

19

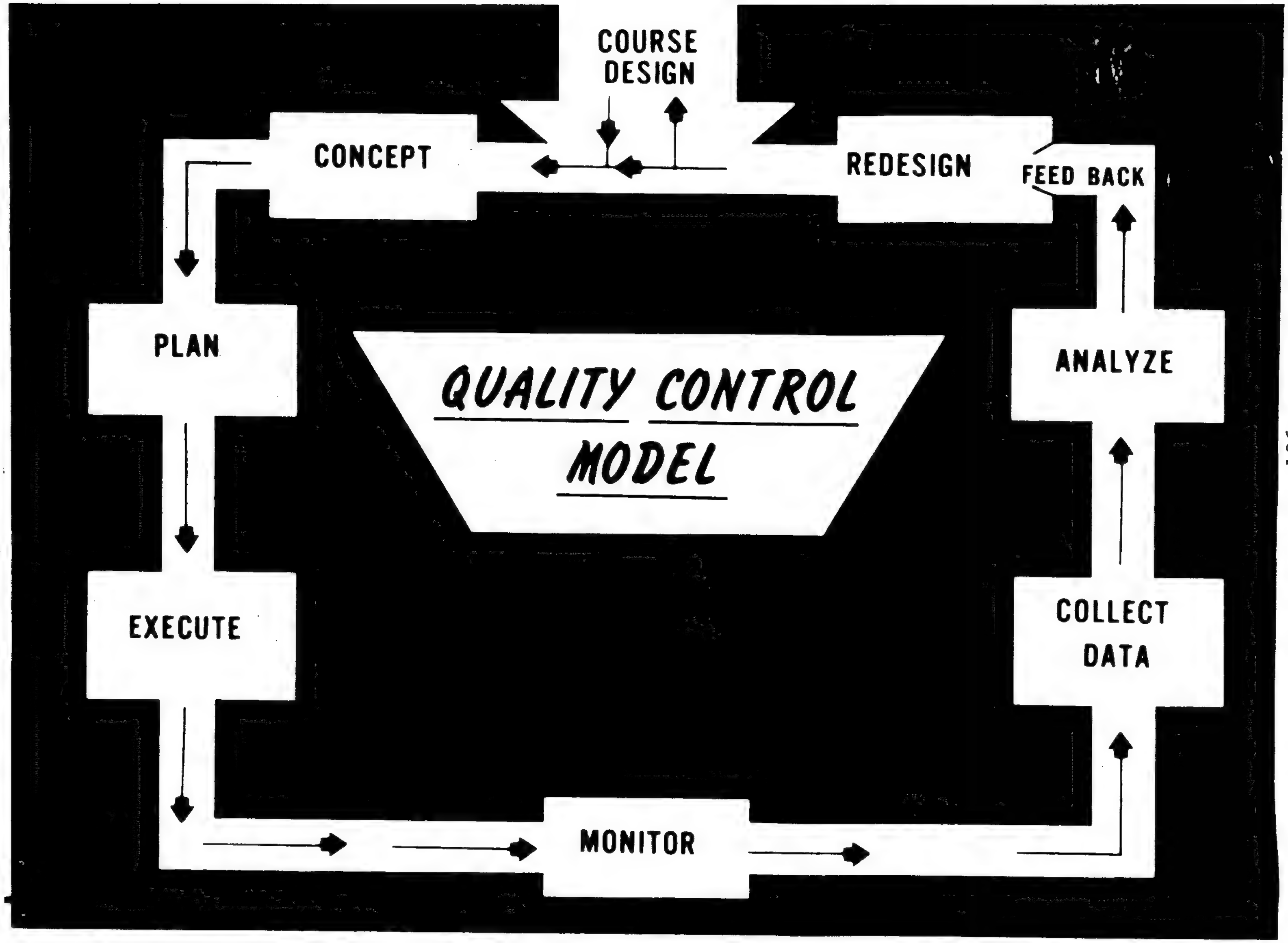
SLIDE #5

SYSTEMS ENGINEERING PROCESS

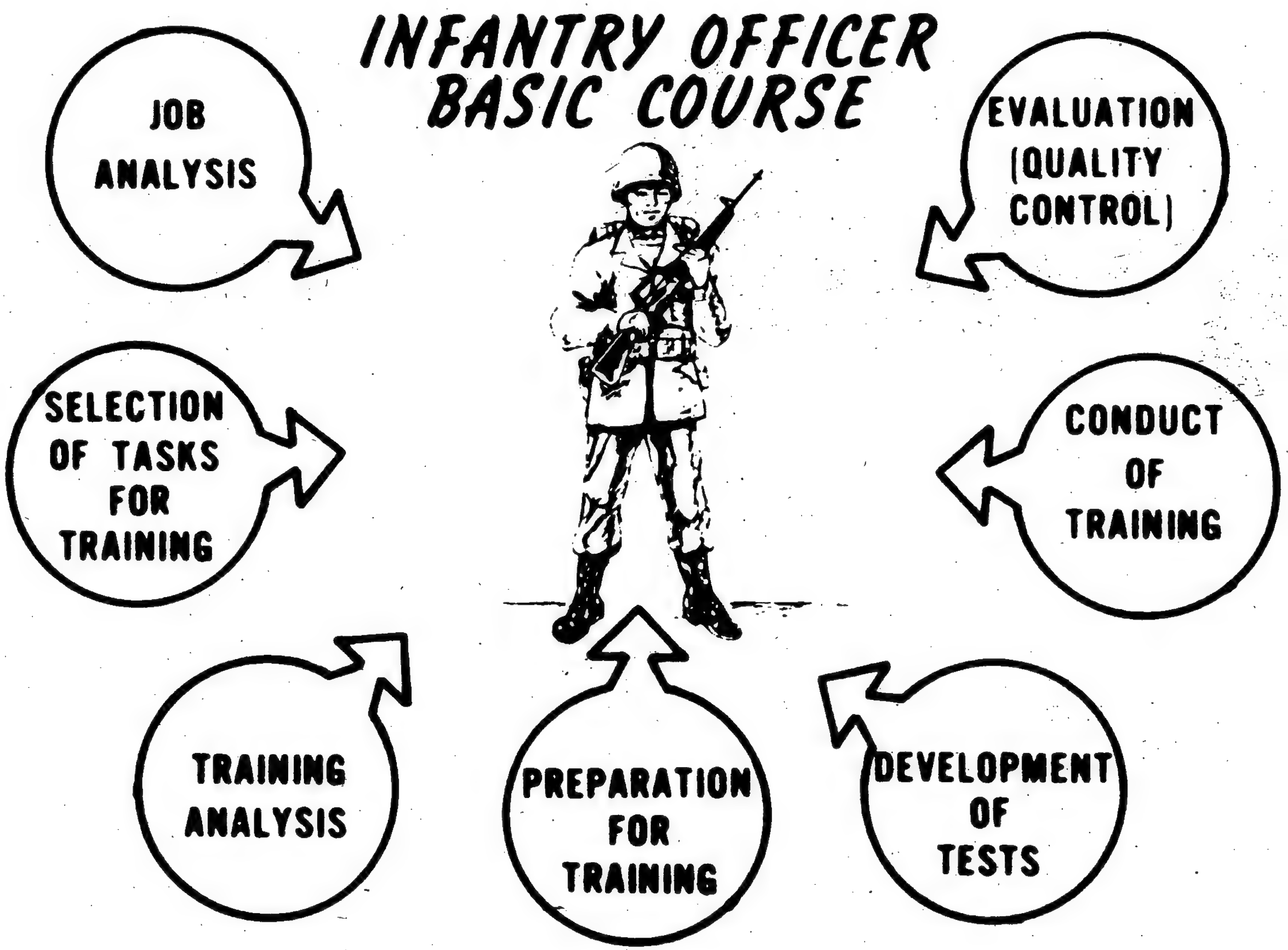


135

SLIDE #8

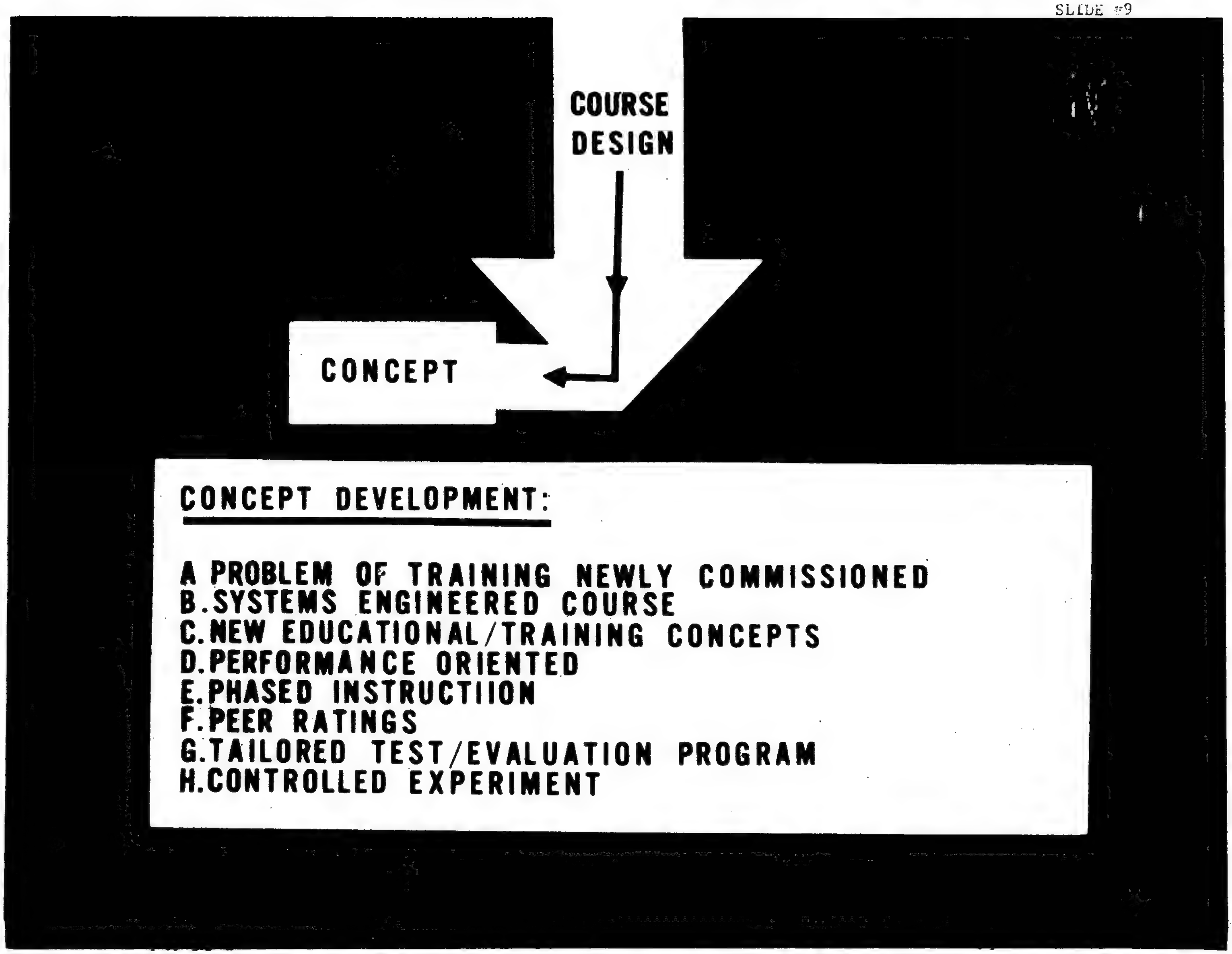

COURSE DESIGN
CONCEPT
REDESIGN
FEED BACK
PLAN
QUALITY CONTROL MODEL
ANALYZE
EXECUTE
COLLECT DATA
MONITOR

SLIDE # 7

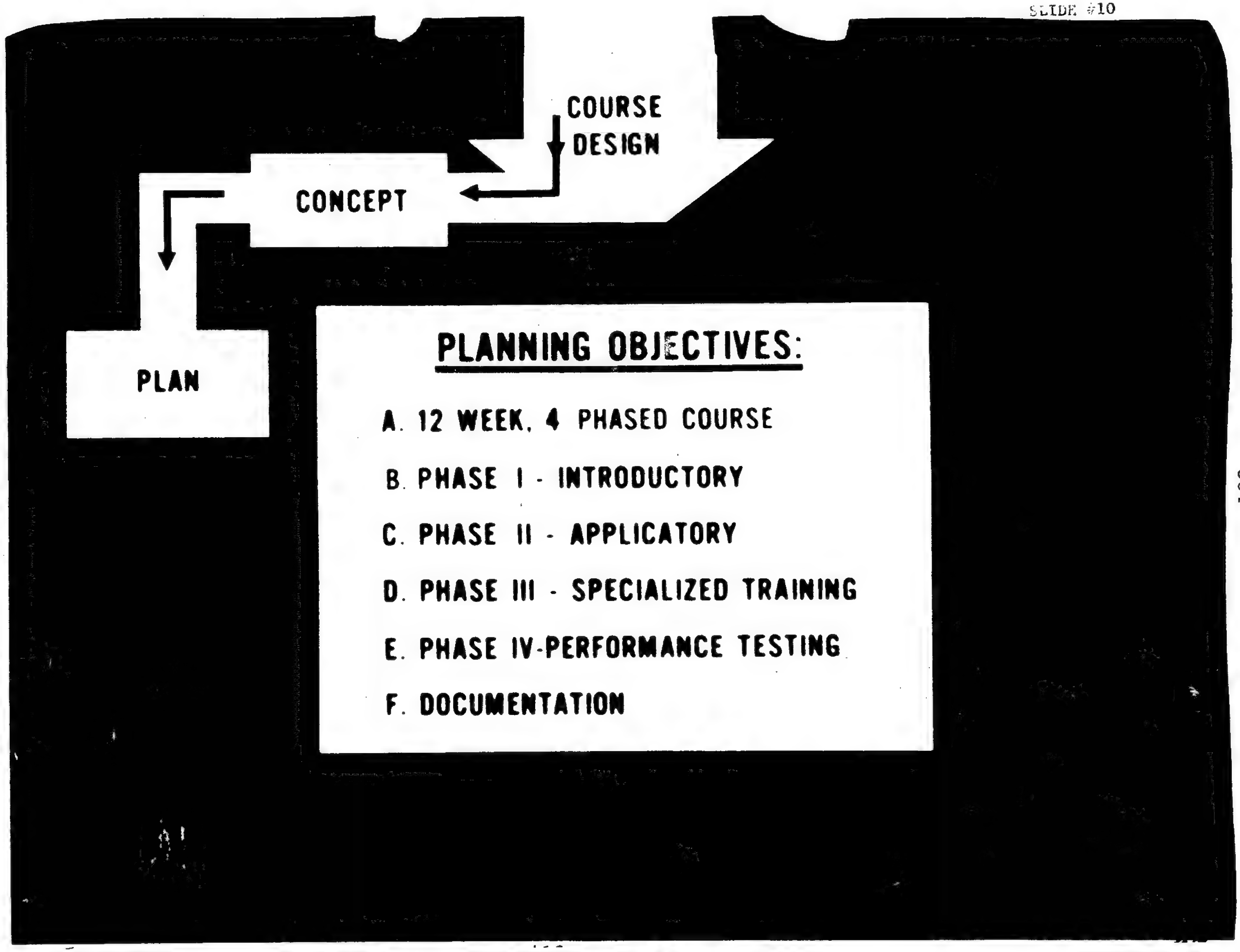


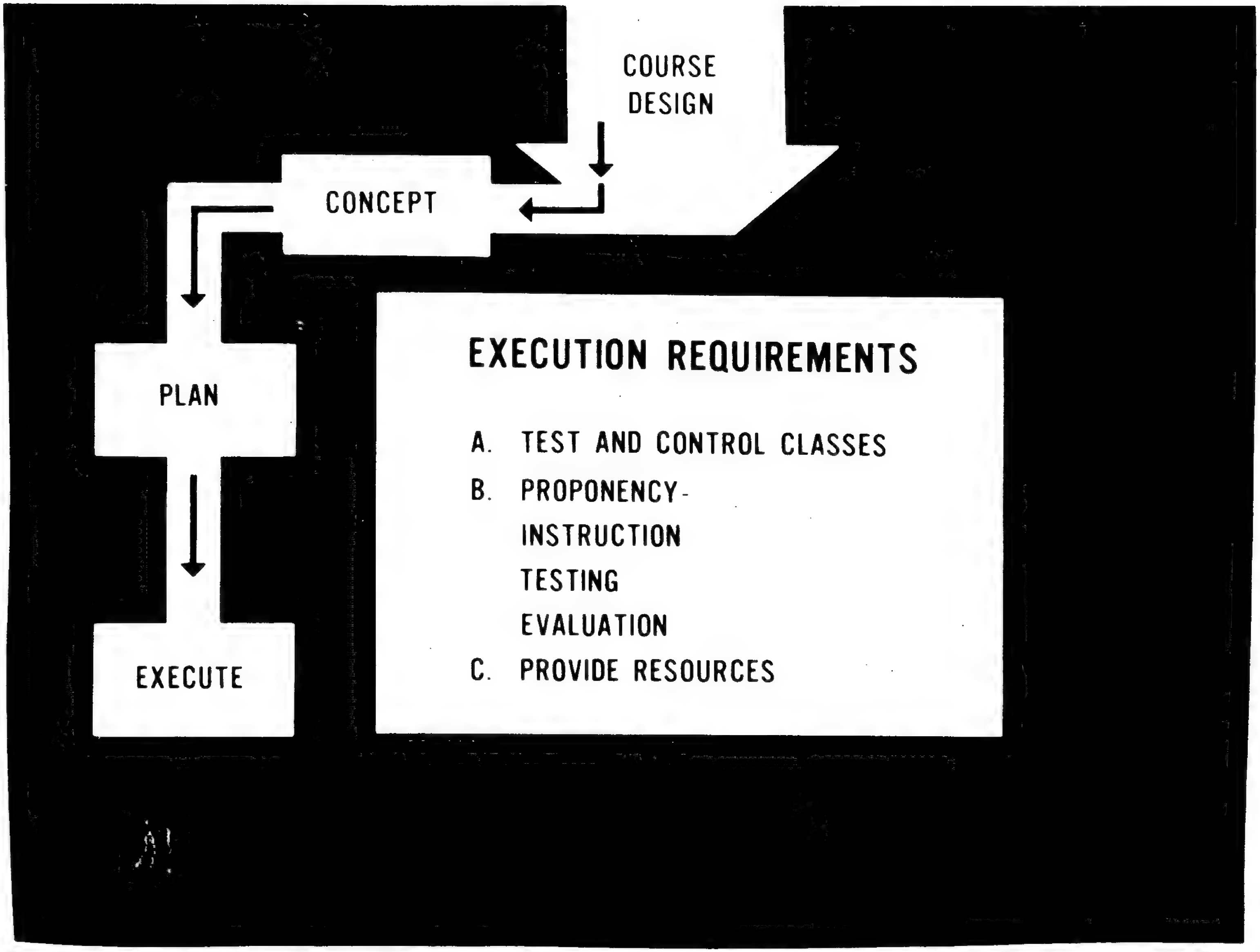
INFANTRY OFFICER
BASIC COURSE
JOB ANALYSIS
SELECTION OF TASKS FOR TRAINING
TRAINING ANALYSIS
PREPARATION FOR TRAINING
DEVELOPMENT OF TESTS
CONDUCT OF TRAINING
EVALUATION (QUALITY CONTROL)

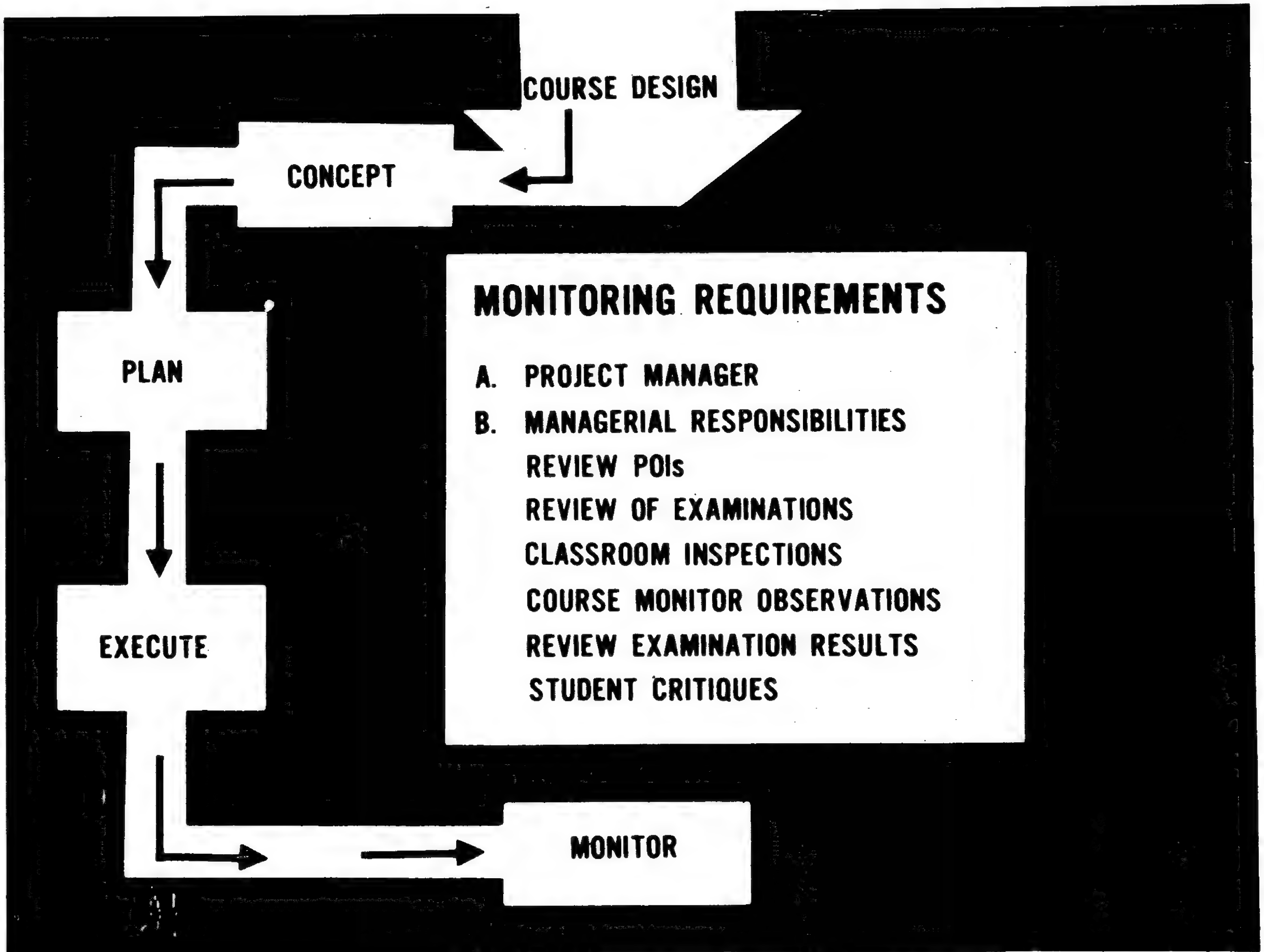
SLIDE #9

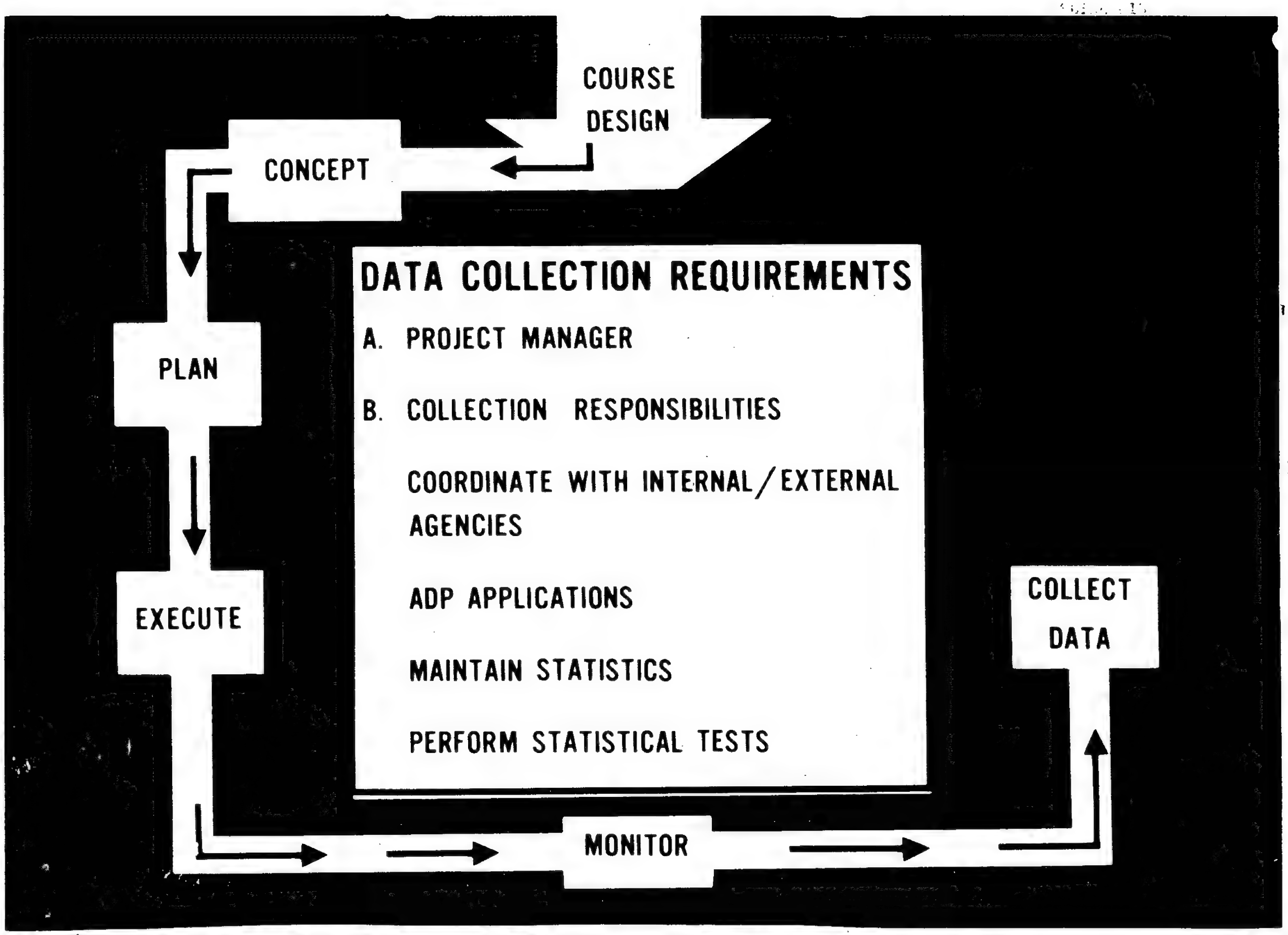

COURSE
DESIGN
CONCEPT
CONCEPT DEVELOPMENT:
A PROBLEM OF TRAINING NEWLY COMMISSIONED
B.SYSTEMS ENGINEERED COURSE
C.NEW EDUCATIONAL/TRAINING CONCEPTS
D.PERFORMANCE ORIENTED
E.PHASED INSTRUCTIION
F.PEER RATINGS
G.TAILORED TEST/EVALUATION PROGRAM
H.CONTROLLED EXPERIMENT

SLIDE #10

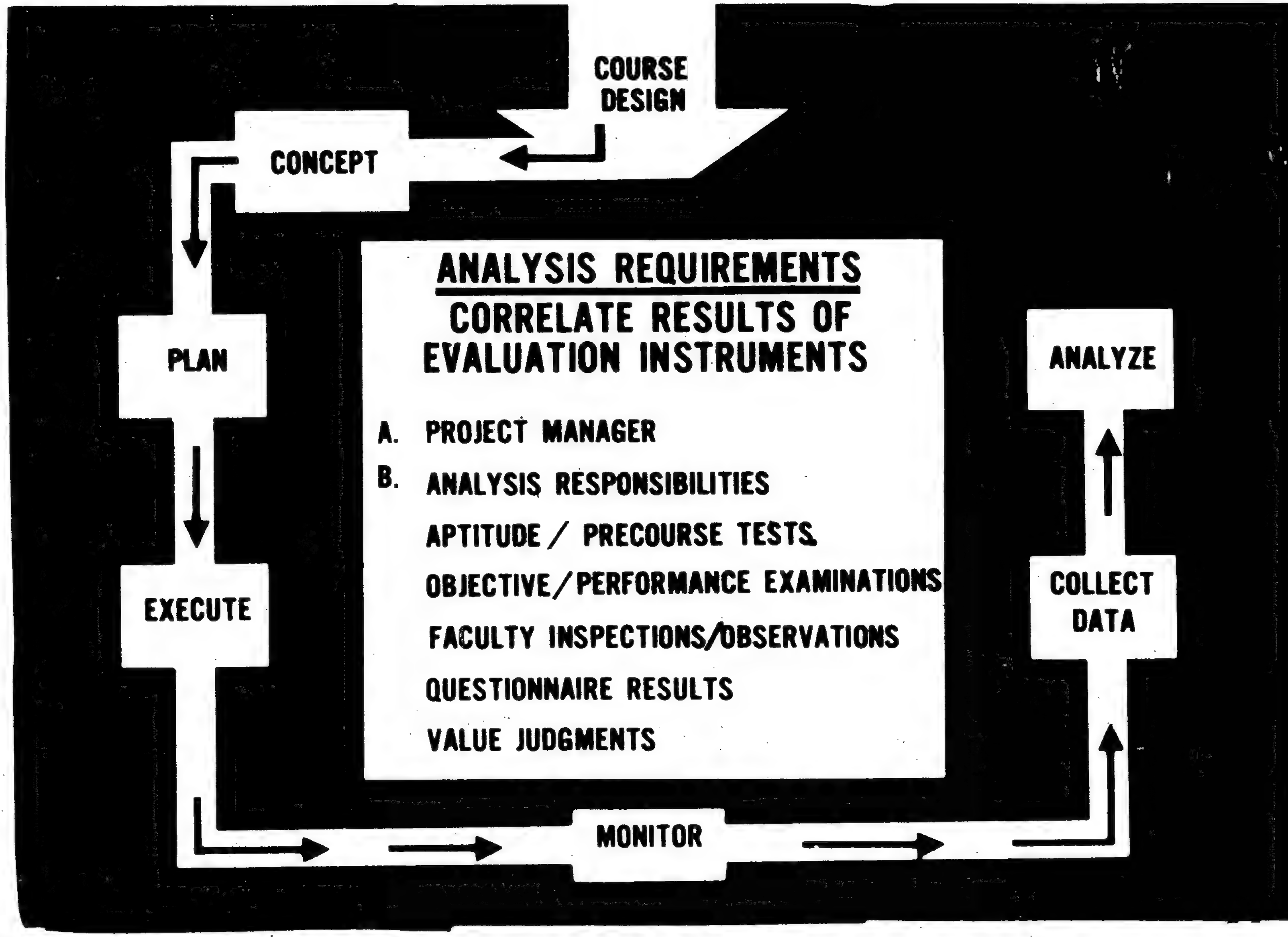

COURSE
DESIGN
CONCEPT
PLAN
PLANNING OBJECTIVES:
A. 12 WEEK, 4 PHASED COURSE
B. PHASE I - INTRODUCTORY
C. PHASE II - APPLICATORY
D. PHASE III - SPECIALIZED TRAINING
E. PHASE IV-PERFORMANCE TESTING
F. DOCUMENTATION


COURSE
DESIGN
CONCEPT
PLAN
EXECUTE
EXECUTION REQUIREMENTS
A. TEST AND CONTROL CLASSES
B. PROPONENCY-
INSTRUCTION
TESTING
EVALUATION
C. PROVIDE RESOURCES

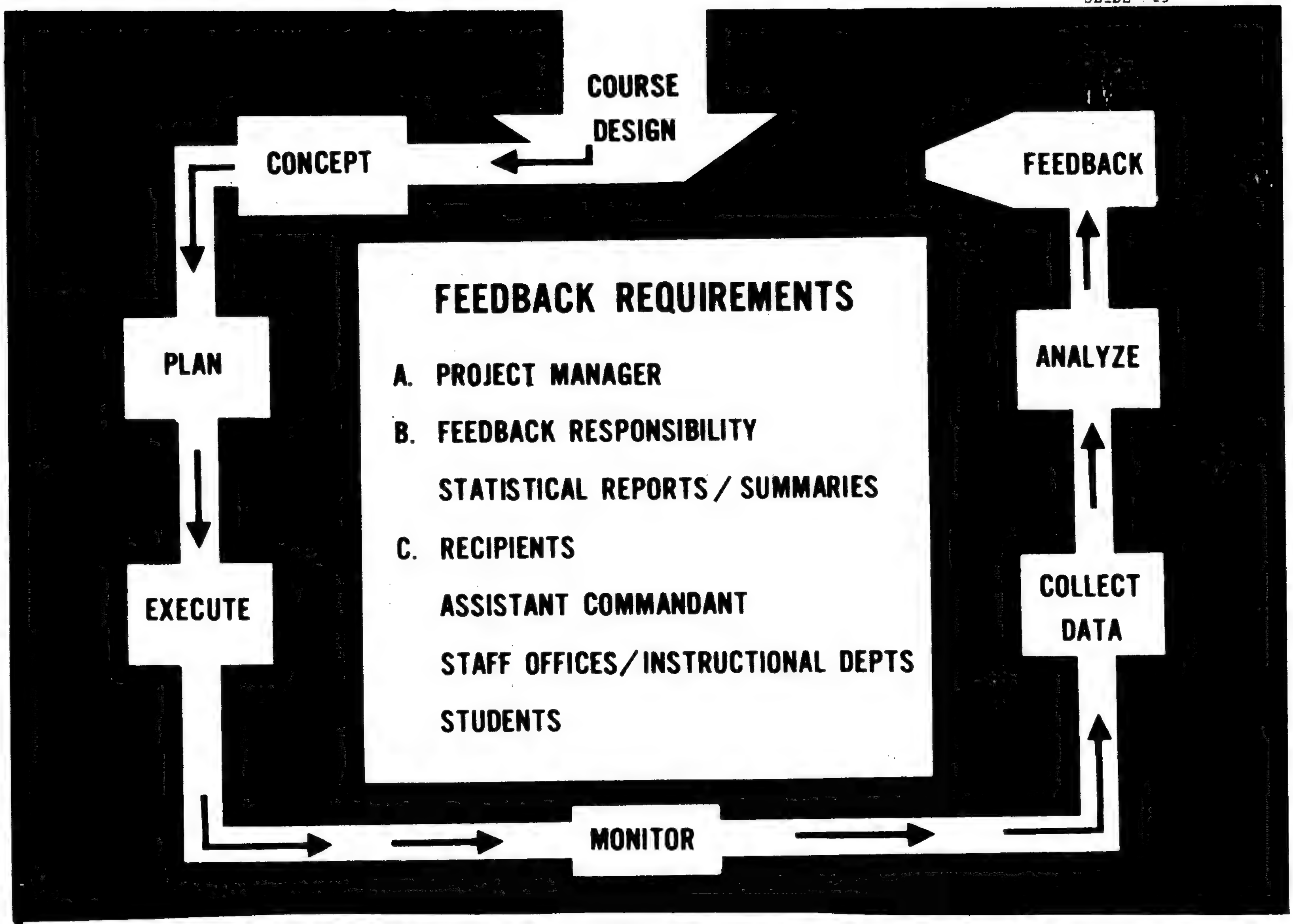

COURSE DESIGN
CONCEPT
PLAN
EXECUTE
MONITOR
MONITORING REQUIREMENTS
A. PROJECT MANAGER
B. MANAGERIAL RESPONSIBILITIES
REVIEW POIs
REVIEW OF EXAMINATIONS
CLASSROOM INSPECTIONS
COURSE MONITOR OBSERVATIONS
REVIEW EXAMINATION RESULTS
STUDENT CRITIQUES


COURSE
DESIGN
CONCEPT
PLAN
EXECUTE
DATA COLLECTION REQUIREMENTS
A. PROJECT MANAGER
B. COLLECTION RESPONSIBILITIES
COORDINATE WITH INTERNAL/EXTERNAL AGENCIES
ADP APPLICATIONS
MAINTAIN STATISTICS
PERFORM STATISTICAL TESTS
MONITOR
COLLECT
DATA

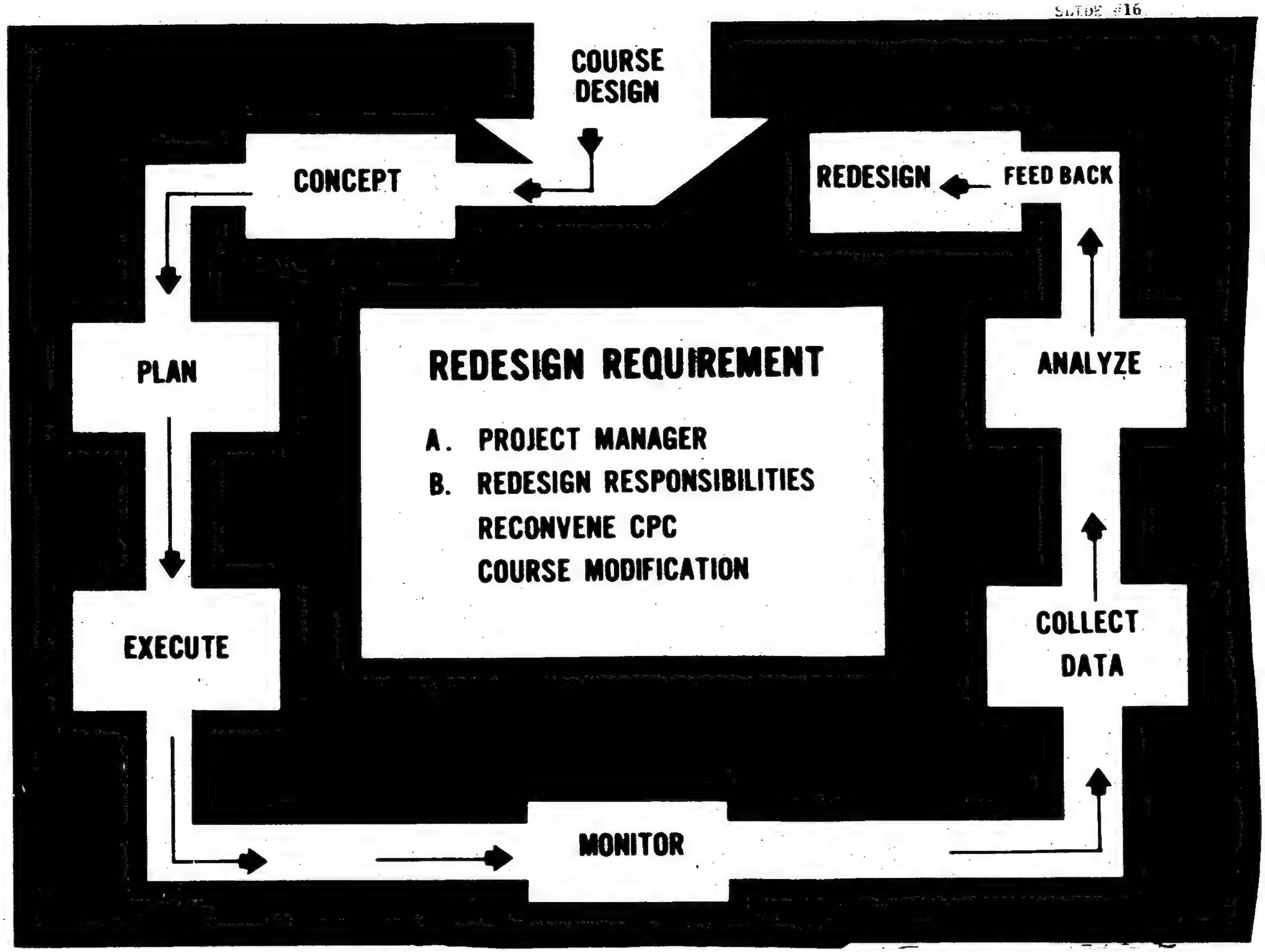
SLIDE #14


COURSE DESIGN
CONCEPT
PLAN
EXECUTE
MONITOR
COLLECT DATA
ANALYZE
ANALYSIS REQUIREMENTS
CORRELATE RESULTS OF EVALUATION INSTRUMENTS
A. PROJECT MANAGER
B. ANALYSIS RESPONSIBILITIES
APTITUDE / PRECOURSE TESTS
OBJECTIVE/PERFORMANCE EXAMINATIONS
FACULTY INSPECTIONS/OBSERVATIONS
QUESTIONNAIRE RESULTS
VALUE JUDGMENTS

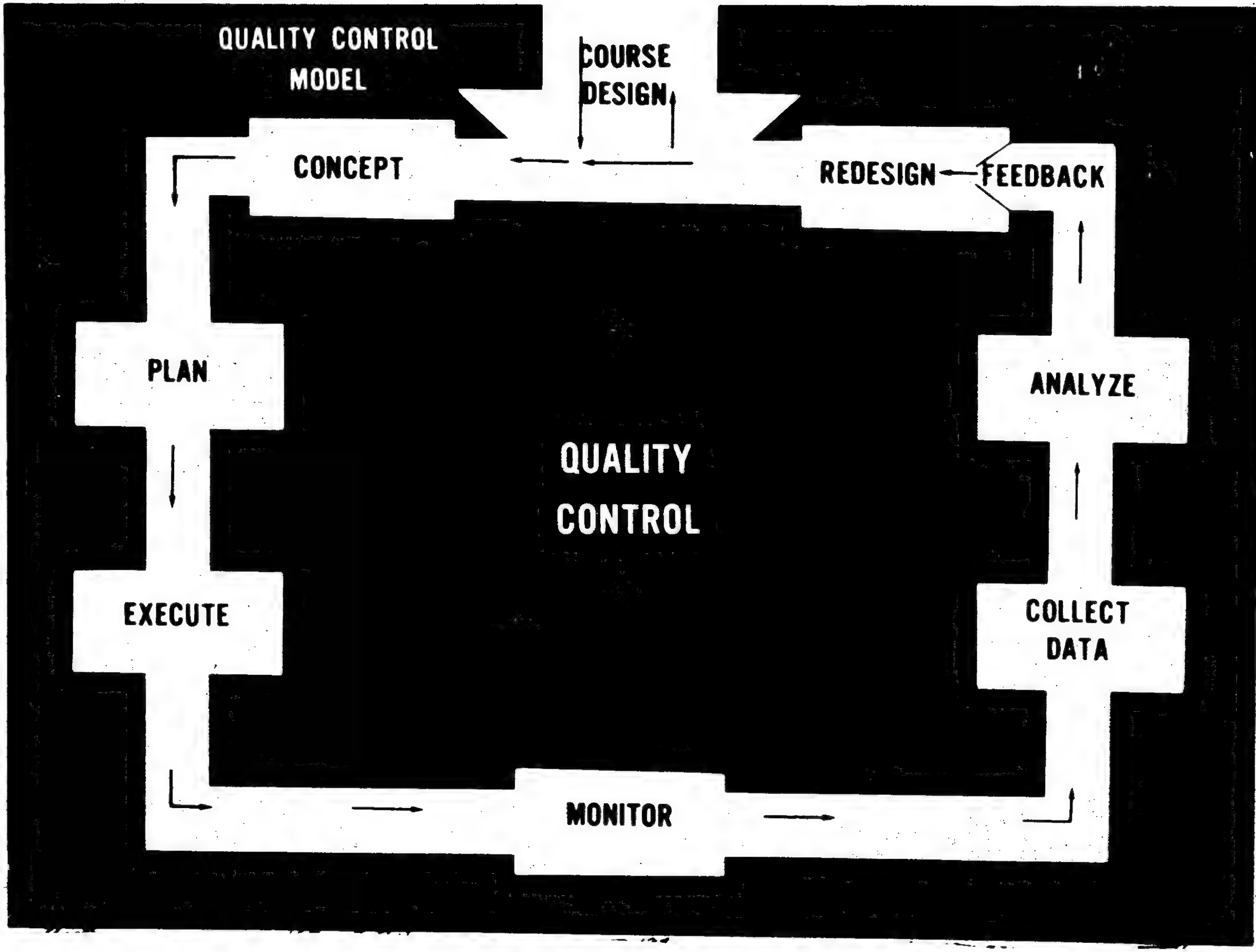
SLIDE #15

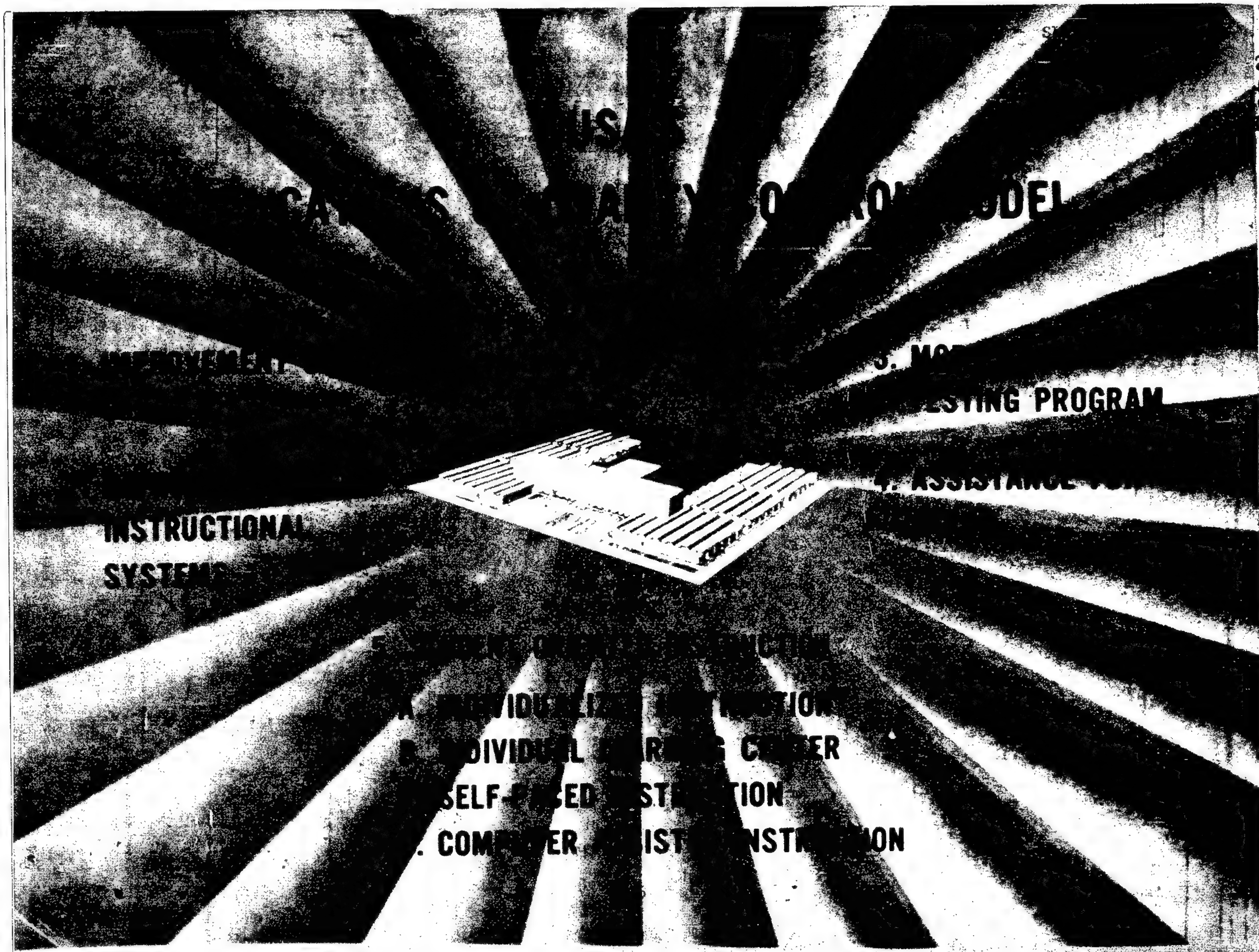

COURSE
DESIGN
CONCEPT
PLAN
EXECUTE
MONITOR
COLLECT
DATA
ANALYZE
FEEDBACK
FEEDBACK REQUIREMENTS
A. PROJECT MANAGER
B. FEEDBACK RESPONSIBILITY
STATISTICAL REPORTS / SUMMARIES
C. RECIPIENTS
ASSISTANT COMMANDANT
STAFF OFFICES/INSTRUCTIONAL DEPTS
STUDENTS

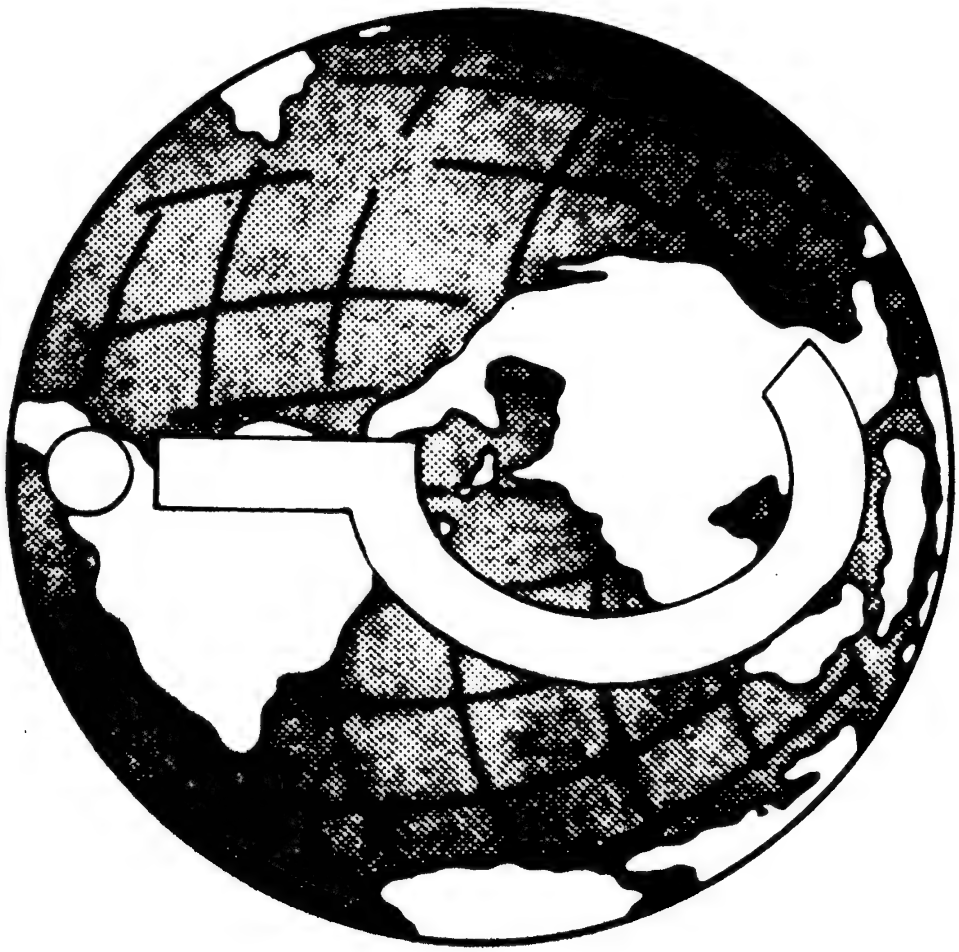
SLIDE #16


COURSE DESIGN
CONCEPT
PLAN
EXECUTE
MONITOR
COLLECT DATA
ANALYZE
FEED BACK
REDESIGN
REDESIGN REQUIREMENT
A. PROJECT MANAGER
B. REDESIGN RESPONSIBILITIES
RECONVENE CPC
COURSE MODIFICATION

SLIDE #17


QUALITY CONTROL
MODEL
COURSE
DESIGN
CONCEPT
REDESIGN
FEEDBACK
PLAN
ANALYZE
QUALITY
CONTROL
EXECUTE
COLLECT
DATA
MONITOR


INSTRUCTIONAL
SYSTEMS
TESTING PROGRAM
B. INDIVIDUAL
SELF-PACED
ION



GENERAL TRAINING SYSTEM (GENTRAS) FIELD EVALUATION ROUTINE

By

JAMES K. MILLER

G-3 Division, Headquarters Marine Corps
Washington, D. C.

GENERAL TRAINING SYSTEM (GENTRAS) FIELD EVALUATION ROUTINE

The General Training System (GENTRAS) field evaluation routine is a computerized program for evaluating formal school course effectiveness. The GENTRAS field evaluation routine is based on the hypothesis that the effectiveness of any course of instruction should be measured through the evaluation of course graduates performing on-the-job by immediate supervisors.

In order to better understand the field evaluation routine and its relationship to GENTRAS, it is necessary to describe GENTRAS and how it works.

Development of GENTRAS, an automated data processing system designed to assist Headquarters Marine Corps in evaluating the qualitative aspects of training, was completed in February 1971. Implementation began in April 1971. It is anticipated that it will take 37 months to fully implement GENTRAS to support enlisted ground training only. Future plans may also see GENTRAS embracing officer and aviation training. A four-man team has been formed to implement and manage the system.

GENTRAS design and support capabilities are based upon the need to ensure that training is directed toward preparing individuals to perform in a job upon completion of training. Heretofore, the Marine Corps has not truly been able to manage the quality of training. The Marine Corps has depended upon the subjective expertise of schools or MOS specialists....or in some cases

tradition....to determine what should be taught and to whom. Hopefully through GENTRAS the Marine Corps will now be able to manage the qualitative aspects of ground enlisted formal training from entry level until retirement.

GENTRAS provides automated support for closing the loop between those who manage training (HQMC), those who conduct training (formal schools) and those who use training (field units) so that training is timely, pertinent and effective. Figure 1 shows the general relationship of GENTRAS to these organizations.

GENTRAS as depicted here includes the Training Management System (TMS) which is physically located at HQMC, operates on their computer (IBM 360-65) uses basically the NMCS Information Processing System (NIPS).

In order that training be appropriate, it is necessary to determine the specific skills required to do the job, then organize courses of instruction such that they provide these skills. Further, to evaluate the effectiveness of training, it is necessary to observe and rate proficiency of graduates insofar as they can or cannot perform in the field. In either event, training requirements are determined based upon validated field performance requirements. See balloon (1) in Figure 1.

<u>GENTRAS Environment and Purpose</u>

Job data is currently gathered by USMC Office of Manpower Utilization (OMU) and analyzed to determine skills and skill

usage required to perform in the job. Based upon the analysis, occupational fields may require restructuring causing new MOS's to be generated and some old MOS's to be deleted. The new structures, however, reflect the actual field job requirements and make evident the changing skill requirements as an individual progresses through his career in the Service.

The OMU task analysis makes available to HQMC, G-3, detailed job descriptions and occupational field structures (1). Course information is provided by the schools (1c). The Training Management System provides automated support for storing raw MOS, course, and field evaluation data (2). Queries against this data base provide the capability for selecting and correlating this information to produce meaningful output (3). Using this information (output) HQMC G-3 can determine apparent deficiencies in training and will, in turn, recommend changes to the schools (4). Schools, acting upon these recommendations, are able to place better trained personnel in the field (5). Through field evaluation of recent graduates, and by resurveying field units, HQMC is able to further evaluate jobs, job structure and training effectiveness, whereupon the cycle may be repeated. Field evaluations are normally conducted on a sample basis; however, they are always conducted for new courses and courses known to be deficient.

Training Management System Files

The initial implementation of TMS requires the use of a number of data sources such as the current MOS Manual, Training

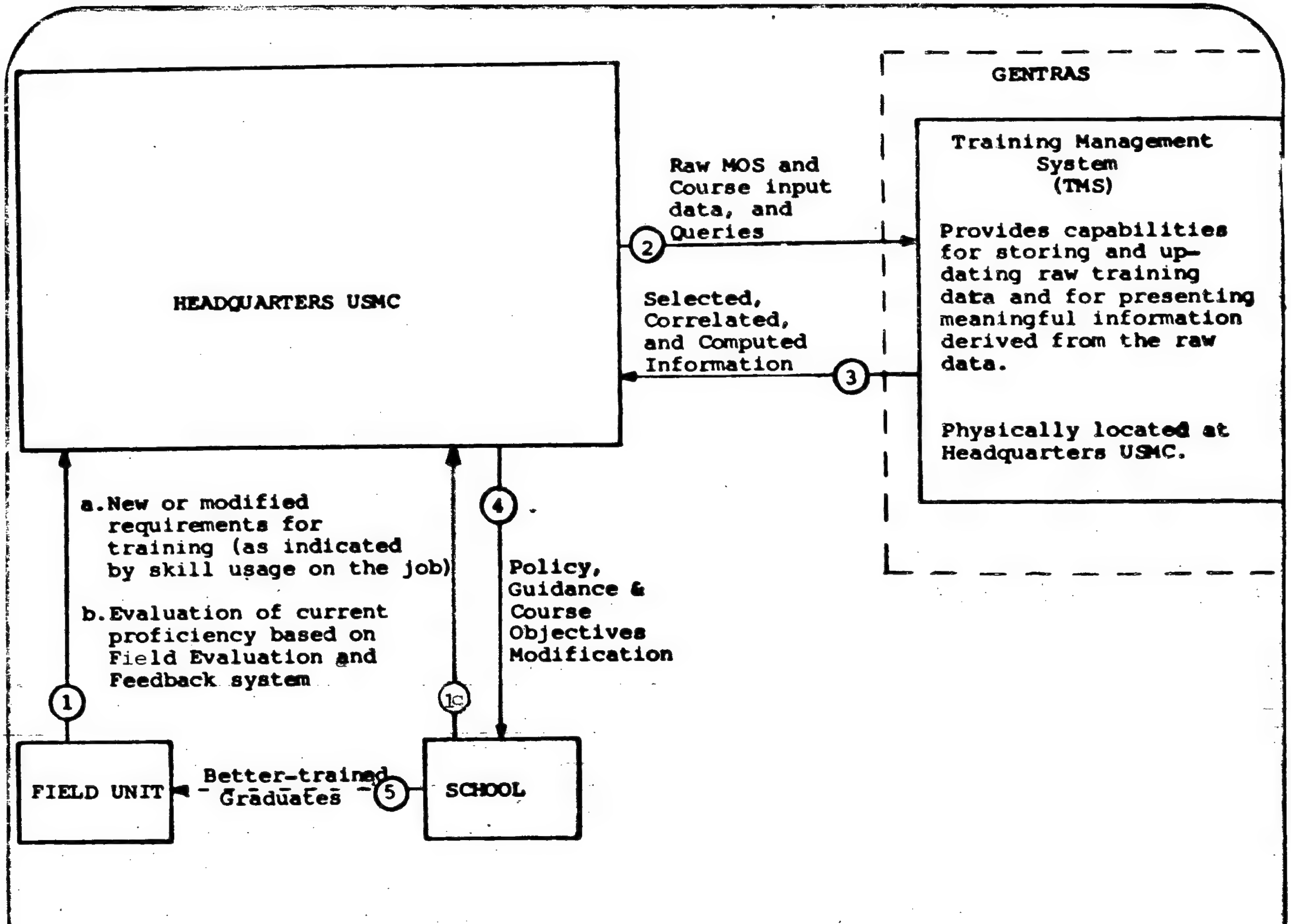


Figure 1. GENTRAS Environment and Purpose

Program Document, Formal Schools Catalog, PRIME reporting system, student records, programs of instruction, as well as previously mentioned task analysis data, and field evaluation data. This raw data is stored in five basic files: Cost, Course, MOS, Trend and Rating. The Cost and Trend files have no bearing on this paper and will not be discussed further. A general description of the three significant files upon which GENTRAS supports training evaluation follows.

1. <u>COURSE File</u>

This file will contain course data -- one record for each course attended by Marines. The data consist of information about the course, classes conducted, course effectiveness, MOS(s) awarded by the course, MOS(s) eligible to attend, courses required prior to attending this course, classification and/or aptitude score requirements for attending this course, skills taught in the course, information about the proficiency with which graduates are able to perform each skill in the field and narrative information about the course.

2. <u>MOS File</u>

This file will contain MOS data -- one record for each MOS. The information consists of the MOS identity, prerequisites for holding the MOS, the range of ranks by skill level of Marines who may hold the MOS, training available to anyone in the MOS and an indication of which courses are required, next MOS(s) that may be awarded in a normal progression through the occupational field, skills performed by persons holding the MOS and skill

usage information, the ID of the course in which each skill is taught and narrative information about the MOS.

3. RATING File

This file will contain data on students -- one record for each student for those classes that are surveyed in the field evaluation process. The information consists of course, class and student ID, student aptitude and classification test scores, performance information in school and in class, skills learned in school and proficiency with which he performed in the field.

Field evaluation data will come from field units. The performance of recent graduates from a surveyed course is rated by their supervisors. They are evaluated approximately 2 to 6 months after graduation from the course and are rated on each skill taught in the course. The exact length of time varies dependent upon course and job complexity. This information is placed in the RATING file along with other information about each student. This information is then processed to determine composite performance of all students rated on each skill. This is called a Field Proficiency Rating (FPR). A low FPR for a skill indicates that it is probably not taught effectively. The average proficiency of each student performing all skills is also determined and a student field rank determined. The field rank is then compared to rank in class. The amount of agreement between these two rankings (expressed in TMS as RANKING) provides some insight into what effect performance in the course had on performance in the field.

RANKING expresses only the amount of agreement between the ranking of students in class against their ranking by field performance; it does not reflect the quantitative difference existing between grades received in class and in the field (disparity). A field called GRADING shows this latter value and is expressed as a percentage of the maximum possible disparity. The product of GRADING and RANKING, expressed as a percentage, and average FPR produce a Training Effectiveness Index (TEI).

Training Management Data Outputs

Queries against the five files produce data outputs in one or more of 18 different standard formats. Output formats are related to up-to-date job and course information, career path information, updated task analysis questionnaire listing, training appropriateness in view of changing job requirements and training effectiveness. Other information can be derived from a combination of the above and obtained by "ad hoc" queries prepared by a NIPS analyst.

Now that the functional aspects of GENTRAS have been described and the relationship that the field evaluation has within GENTRAS, a description of the field evaluation routine will be meaningful.

Field Evaluation of Course Graduates

The diagram in Figure 2 represents the flow of information associated with the field evaluation of recent graduates. The information on the graduating class, students and grades (1) is sent to HQMC and provides the basis for field evaluations discussed earlier. Since field evaluation will only be performed

on a sample basis or on new courses and courses known to be deficient, the detailed student information contained in the RATINGS file need not be passed to HQMC as a matter of routine; however, the class number, start and end dates and the number of students starting and ending each class must be sent to HQMC. This information is recorded in the COURSE file and totals provided for input to the COST file. The decision to do a field evaluation on a class of graduates (2) should be made at an early enough date so that steps 3 through 6 can be completed prior to graduation. The COURSE file will be queried (3) to produce a listing of the skills taught in the course (4). The appropriate instructions are then attached and the questionnaire and answer booklet (See attached Commandant of the Marine Corps letter) are reproduced (5) and sent to the schools for insertion into the student's personnel jacket (6) prior to his leaving school for assignment to a field unit (7).

At the designated time after graduation, each graduate's supervisor will complete the field evaluation questionnaire and return it to HQMC, G-3 (8). G-3 will obtain the graduates' classification and aptitude test scores from HQMC Personnel (9) and code all data for the graduating class (10). Graduate data and questionnaire responses will be provided in coded form (11) to Data Systems for keypunching (12) and input to the RATING file where it will undergo processing (13) as mentioned earlier.

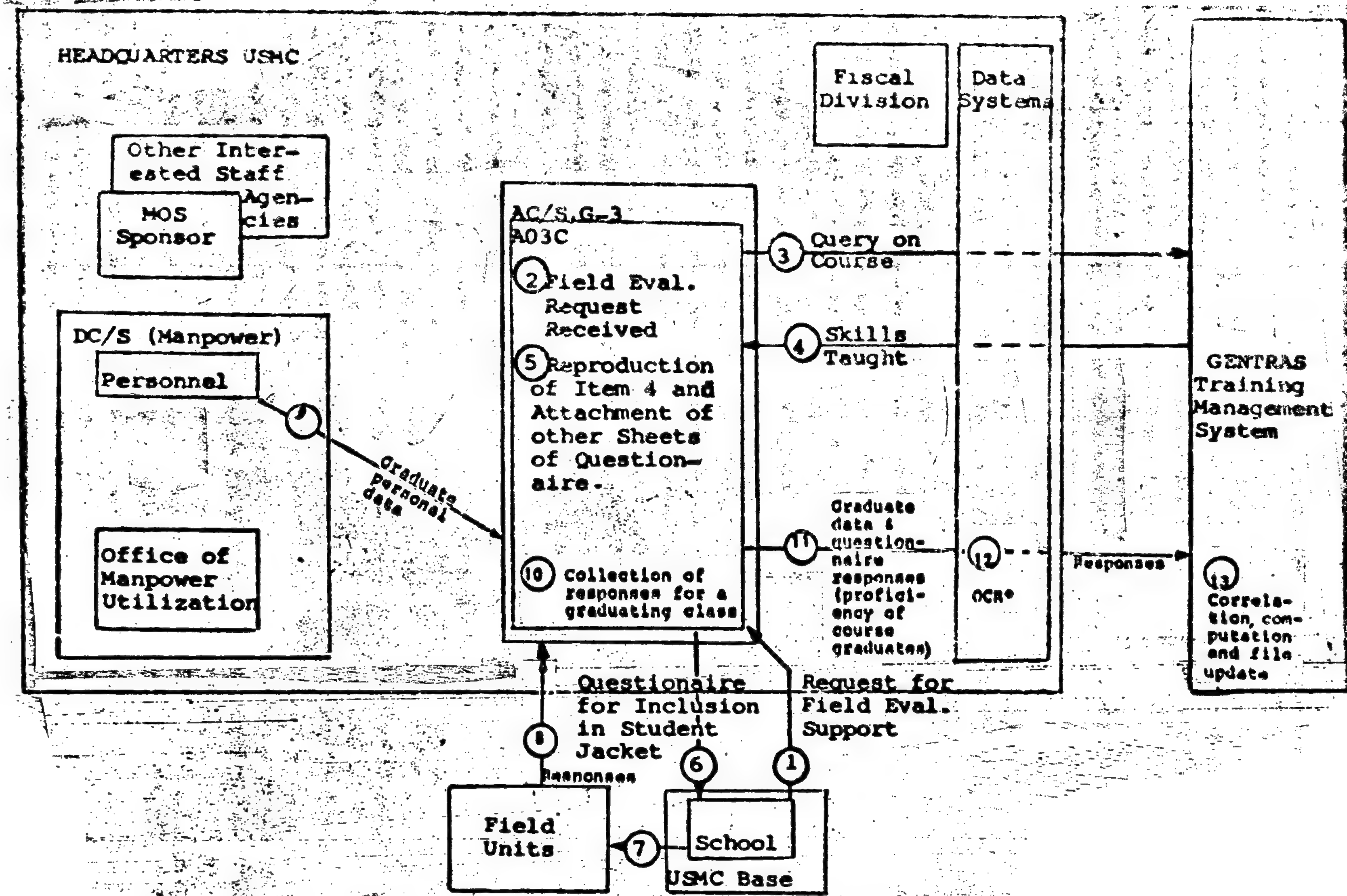


Figure 2. Field Evaluation of the Proficiency of Course Graduates

The Training Management System supports the field evaluation routine by processing the returned questionnaires to determine:

- Average performance ability of the class for each skill taught in the course.

- Overall field performance ability of individual graduates and their ranking within the class of graduates in field performance.

- Correlation between students rank order in class and rank order in field performance.

- Quantitative difference between course grades and field performance ratings. (It is possible that a high correlation exists between the rankings; yet, a significant quantitative difference may exist between the grades and the average ratings for the students.)

- Who is rating higher -- instructor of field supervisor and by how much?

- Overall Training Effectiveness Index (TEI) for the course. (TEI combines correlation, amount of difference between grades and performance, and average field performance of the composite class.)

- Training required to support field performance that is not provided by a course of instruction.

- Training provided by a course of instruction which is not required.

- Past performance ratings and trends toward extremely high or low performance for each skill and course evaluated.

Once analyzed, evaluated and validated at HQMC and confirmed with field units the above information can result in the following type action.

- deletion of course material
- inclusion of new subject matter
- reduction of emphasis
- increased emphasis
- course evaluation (testing methods and instruments) in need of modification.
- improvement of instructor quality

The ultimate result is that the Marine Corps is capable of managing the quality of training to ensure that it is timely, pertinent and effective.

Summary

Before appropriate training can be conducted it is necessary to identify the skills required to perform in a specific job (MOS). It is then necessary to design courses of instruction so that they train personnel to perform the skills. Finally, field supervisors must observe and rate the proficiency of school graduates in the performance of their jobs. This is done through the field evaluation routine and is considered the "truth teller" of the General Training System. Field evaluation will confirm training effectiveness and validate that the Marines are teaching what should be taught.

DEPARTMENT OF THE NAVY
HEADQUARTERS UNITED STATES MARINE CORPS
WASHINGTON, D.C. 20380

IN REPLY REFER TO
A03C53-rlc

From: Commandant of the Marine Corps
To: Commanding Officer of

Subj: General Training System (GENTRAS) Field Evaluation

Ref: (a) MCO P1500.12D

Encl: (1) GENTRAS Field Evaluation Questionnaire

1. Reference (a) requires field commanders to feed back information to schools relative to the appropriateness and effectiveness of the training received by course graduates. The General Training System (GENTRAS), currently being implemented, provides an automated support in this area.

2. The individuals best able to judge the quality of training provided by schools are the immediate supervisors of recent graduates since they observe the performance ability of recent graduates firsthand.

3. Enclosure (1) is the GENTRAS field evaluation questionnaire to be used for rating recent graduates' performance. It consists of three parts:

a. Part I provides the instructions for completing the form.

b. Part II is the ANSWER SHEET to be used for recording the performance ability of the individual on the skills taught in the course.

c. Part III is a list of the skills taught in the course.

4. It is requested that the immediate supervisor of the above named Marine complete Part II of enclosure (1) between the dates specified thereon and return it to the Commandant of the Marine Corps (Code A03C) not later than 31 August 1971.

Subj: General Training System (GENTRAS) Field Evaluation

5. Every attempt has been made to keep the automated feed-back evaluation process as simple as possible while retaining the capability to measure training effectiveness on an objective basis. Comments are invited in this regard and may be included in the "COMMENTS" area of the "ANSWER SHEET."

C. B. Drake

C. B. DRAKE
By direction

GENERAL TRAINING SYSTEM

(GENTRAS)

FIELD EVALUATION

QUESTIONNAIRE

ENCLOSURE (1)

PART I

INTRODUCTION

You, as the supervisor of one or more recent graduates of a formal school, have been selected to participate in an evaluation of the training received by them. Please be as objective as possible.

This is an evaluation of the training received by the individual being evaluated by you. It is not a fitness report nor will the information provided by you be used for any purpose other than to improve training quality thereby providing individual Marines to field units who are better able to perform billet job assignments.

The attached list of skills reflects those taught in the school attended by the individual being rated. It may not be a complete list of the skills needed by the individual to perform in the assigned billet. If this is the case, your attention is directed to the COMMENTS area of the questionnaire ANSWER SHEET.

This is not a test. Neither you, the individual that you are rating nor your unit will be marked or judged, in any way, on the information which you provide. Your individual answers will be combined with the answers of other supervisors in order to determine how effectively training satisfies field requirements.

GENERAL INSTRUCTIONS

1. In the event that the individual is performing in an MOS other than the one for which he was trained, please complete as many ratings as possible. As a minimum, complete the BILLET MOS entry on the ANSWER SHEET.

2. Record your ratings of the individual's performance on the ANSWER SHEET(s) provided. Use the most appropriate number from the chart shown on the following page for each skill listed on PART III of this questionnaire.

3. You are to rate the performance ability of the individual at essentially the level it was when he joined your unit from school. Try to eliminate the influence that remedial OJT may have had on his current performance.

4. If you are undecided between two ratings, choose the one closer to 1 or closer to 7. e.g., if you are undecided between a rating of 2 and 3 for a skill, choose 2 since 2 is closer to 1. If you are undecided between a 5 and 6, choose 6 since it is closer to 7.

5. If the individual has not performed a skill since joining your unit, enter a zero.

6. The ANSWER SHEET provides space for your comments. Please feel free to make any comment. Give special attention to skills needed by Marines that are not included in the list; and skills included in the list that are not needed. If you feel strongly that individuals are not being trained adequately in some aspect of their job assignment, please note it in this area. All pertinent comments will be tabulated and included in recommendations made to improve training.

7. It is requested that this questionnaire be completed between 16 August 1971 and 27 August 1971 and forwarded to the Commandant of the Marine Corps (Code AO3C).

ENCLOSURE (1)

FIELD PERFORMANCE RATING (FPR) ASSIGNMENT CHART

Choose appropriate value From this column for each skill and enter in RATING column on ANSWER SHEET

Use the following three columns to determine the RATING Value for the individual's performance on each skill on the skill list.

QUALITY PRODUCED	SUPERVISION REQUIRED	TIME REQUIRED	OVERALL PERFORMANCE	RATING IS
Individual has not performed this skill since assignment to unit				0
None; unable to perform	Constant; required total training	Very Excessive	Unacceptable	1
Unacceptable	Close	Excessive	Poor	2
Much Rework Required	Some	Somewhat Excessive	Below Average	3
Little Rework Required	Usually None	Within reason most of the time	Average	4
No Rework Required	None	Always within reason	Above Average	5
Exceeds Normal Quality	None	Less than normally required time	Excellent	6
Exceeds Normal Quality	None, and in addition, assists others	Always Ahead of Schedule	Outstanding	7

PART II

ANSWER SHEET

Do Not Use
Course SSC:______
Class No.:______
POI Rev No.:______

Date Form Completed:____________________
Students SS Number :____________________
Current Billet MOS :____________________

Choose the most appropriate performance rating for each skill on the attached list. Consult PART I for full definition and guidelines for choosing appropriate rating value. NOTE: Double check to make sure that you maintain proper question-answer relationship.

Skill Number	Rating
001	______
002	______
003	______
004	______
005	______
006	______
007	______
008	______
009	______
010	______
012	______
013	______
014	______
015	______
016	______
017	______
018	______

Skill Number	Rating
019	______
020	______
021	______
022	______
023	______
024	______
025	______
026	______
027	______
028	______
029	______
030	______
031	______
032	______
033	______
034	______
035	______

Skill Number	Rating
036	______
037	______
038	______
039	______
040	______
041	______
042	______
043	______
044	______
045	______
046	______
047	______

COMMENTS: (Attach additional sheets as necessary)____________________

PART III

PLEASE VERIFY THE SKILL NUMBER WITH THAT PRINTED ON THE ANSWER SHEET

COURSE TITLE: UNIT DIARY CLERK

SSC AND SUFFIX: 01E

POI REVISION NUMBER: 2

	* * * * * S K I L L S	T A U G H T * * * * *	
SKILL NUMBER	ACTION	MODIFIER	OBJECT
001	AUDIT		MPR(S)
002	AUDIT		OPR(S)
003	AUDIT		OQR(S)
004	AUDIT		SRB(S)
005	AUDIT		UTR(S)
006	AUDIT	INCOMING	UNIT DIARIES
007	AUDIT	OCR	FORMS
008	AUDIT	PAY INFORMATION	TRANSCRIPT
009	AUDIT	VISUAL	AUDIT SHEETS
010	COMPARE	COMPUTER PRTOUTS	SOURCE DOCS
011	COMPUTE		LEAVE
012	COMPUTE	PROCEED/DELAY	TRAVEL
013	COMPUTE	TIME	LOG
014	CORRECT	OCR	FORMS
015	DRAFT		REPORTS
016	DRAFT		UNIT DIARY
017	DRAFT	DATA TRANSCRPTN	FORMS
018	DRAFT	MILITARY	PAY FORMS
019	DRAFT	MISCELLANEOUS	CORRESPONDENCE
020	DRAFT	PERSONNEL	FORMS
021	FILE	PERSONNEL	PLATES
022	MAINTAIN		LOGS
023	MAINTAIN	FILE ON	UNIT DIARY
024	MAINTAIN	INDIV RECORD	CARDS
025	MAINTAIN	JOIN/DROP	FILE
026	MAINTAIN	PERSONNEL	STATUS BOARDS
027	MAINTAIN	RECORDS	FILE
028	MAKE		DOG TAGS
029	MAKE	PERSONNEL	PLATES
030	MAKE	PERSONNEL	ROSTERS
031	PREPARE	IDENTIFICATION	CARDS/FORMS
032	PROVIDE	MMS PROCS	TECH ASSISTANCE
033	REVIEW		UTR(S)
034	REVIEW		WORK
035	RESEARCH	REF(S) FOR	UNIT DIARY
036	TRANSCRIBE	UNIT DIARY ENT	PAS
037	TYPE		CERTIFICATES
038	TYPE		MATS
039	TYPE		REPORTS
040	TYPE		STENCILS
041	TYPE		UNIT DIARY
042	TYPE		WARRANTS
043	TYPE	DATA TRANSCRPTN	FORMS
044	TYPE	MILITARY PAY	FORMS
045	TYPE	PERSONNEL	FORMS
046	TYPE	PUNITIVE	FORMS
047	UPDATE	PERSONNEL	PLATES

DEVELOPMENT OF THE
NAVY ADVISOR PROFILE REPORT

by

Ted M. I. Yellen

Personnel Measurement Research Division

Naval Personnel Research and Development Laboratory
Washington, D. C.

The evaluation instrument described in this paper is designed to assess an individual's potential suitability for a Vietnam advisory assignment. Coming, as it does, at a time when our national effort is directed toward pulling American forces out of Vietnam, it may appear inappropriate to be concerned with improving the selection of naval advisors to Vietnam. However, barring a major change in U. S. policy, the timetable for American withdrawal depends to a large extent on the continuing progress of Vietnamization.

The Navy's part in the Vietnamization program has involved turning over more and more of the primary combat responsibility to the Vietnamese, along with many of the naval vessels and other military equipment needed to implement that responsibility. As the Vietnamese receive U. S. naval equipment, more U. S. naval advisory personnel are needed. The advisor assists the Vietnamese in planning and executing naval operations, and provides guidance and advice on various non-military, as well as military, matters. Since he is often the only American assigned to an all-Vietnamese naval unit, his effectiveness as an advisor depends not only upon his good judgment and expertise, but also upon his capacity to function alone in an unfamiliar setting with people whose customs and language are different from his own.

It goes without saying, therefore, that an ineffective advisor contributes nothing to the Vietnamese capability to operate their own Navy; and the poor advisor can sometimes even have a seriously adverse effect on the advisory mission. It is clear, then, that the overall success of the advisory mission depends on the careful selection of individuals for advisory assignment.

Before the Navy Advisor Profile Report was incorporated into the present selection procedure, advisor selection was primarily based on the assumption that a person who had performed successfully in the past would probably perform equally well in the future. The trouble with the application of this logic to advisor selection is that the advisory role encompasses a range of conditions which differ noticeably from the

usual naval duty assignments. In Vietnam, the advisor can expect to find different social and ethical values, an unfamiliar language, limitations on his authority and freedom of action, poverty, inadequate housing, isolation, and hostile actions. Because of these different conditions, evaluation based on previous performance proved a poor predictor of effective advisory performance. While most advisors have excelled in their previous duty assignments, some of them prove unable to adjust to the necessary changes and, therefore, perform notably less effectively in the advisory assignments.

Investigation into the advisor selection problem has shown that it is usually the human problems associated with working in a different culture that are likely to be critical to the success or failure of the advisor. At present, the Navy's Personal Response Program acts as a training device to improve the prospective advisor's ability to function more effectively in a different culture. However, because of the difficulty involved in changing one's personality and character in order to adjust to the life and work environment found in Vietnam, the training program alone does not satisfy the need for effective advisory performance. A specific advisory selection program is prerequisite to enhancing the Navy's advisory mission in Vietnam.

In order to understand and analyze effective advisory performance, it was necessary to identify behavioral factors that relate to the culture and advisory role in Vietnam. From a variety of sources, such as personal interviews with former Vietnam advisors, examination of relevant literature, and discussions with consultants experiences in selecting and training Americans for foreign assignments, the following 12 personal characteristics were identified as required behavioral factors for Vietnam advisors to possess.

1. Patience
2. Tact, diplomacy, social skill
3. Friendliness, sense of humor, sociability
4. Persistence, perseverance
5. Adaptability
6. Self-reliance, resourcefulness, ingenious
7. Empathy
8. Leadership and organizational ability
9. Emotional stability
10. Instructional ability
11. High moral standards, incorruptible
12. Job dedication, motivation

Fig. 1.--12 Behavioral Factors Found Necessary For Effective Advisor Performance

In discussions with former Vietnam advisors, it was found that the meaning of behavioral factors in the normal military setting was quite different from the meaning attached to it in the Vietnam environment. As a result, the factor titles per se were not adequate descriptors of behavior and, therefore, did not provide adequate information for personnel assessment. Because of this, it was necessary to obtain real-life descriptions of work experiences as they relate to the behavioral factors. For example, factor "Patience" would be defined by what patience means in a Vietnam setting.

In order to define behavioral factors specific for effective advisor performance, a questionnaire was developed and administered to a selected sample of officers and petty officers who were serving in advisory billets and also to a selected number of personnel who had returned from advisory assignments. In the sample, 200 advisors were asked to describe real-life situations in Vietnam relating to the behavioral traits deemed paramount to successful advisor performance. The advisors were also asked to give examples of good and poor behavior for the individual behavioral traits. After the descriptions were collected, they were abstracted and categorized to form a composite picture of behavioral essentials. These categories then formed a behaviorally based starting point for developing operational definitions of performance behaviors regarded as crucial to advisor effectiveness. In addition to defining behavioral factors, the advisors were also requested to list the most critical personal qualities to look for in selecting officer and enlisted personnel for Vietnam advisory assignments. The most frequently mentioned qualities, which do not duplicate those already presented in Fig. 1, are presented in Fig. 2.

1. Technically proficient
2. Willing to listen and learn
3. Good at handling people
4. Racially non-prejudiced
5. Jack-of-all-trades
6. Mature in judgment and actions
7. Able to take care of himself
8. Uses common sense
9. Well-rounded Navy knowledge
10. Has pride in appearance
11. Performs well without supervision
12. Absence of superior attitude

Fig. 2.--12 General Qualities Found Necessary For Effective Advisor Performance

An experimental evaluation form, with instructions and behavioral definitions, was developed and field tested with personnel stationed at Norfolk Naval Station. Based upon the field test results the Navy Advisor Profile Report (NAPR) shown in Fig. 3 and instruction manual were developed and are currently in fleet-wide operational use. When a Navyman requests advisory assignment, he is rated on this form by his supervisor. The supervisor forwards the completed form to the Bureau of Naval Personnel where Vietnam detailers consider the applicant for advisory training and eventual assignment.

As you can see, the form consists of two major evaluation sections, one pertaining to predicting future performance, and the other pertaining to observed past performance. The prediction of future performance is based on these 12 behavioral factors. The definitions for each factor are contained in the manual. For example, Fig. 4 contains the definition for the factor "Patience & Persistence." Below the definition is a 7 point rating scale with points A, D, and G accompanied by descriptive statements. The evaluator rates the applicant on the factor by selecting the scale value that he feels would best predict his behavior in a Vietnam advisory assignment.

In the General Qualities section of the NAPR, the evaluator simply circles the scale value which he feels best describes the individual. For example, if he feels the applicant is technically proficient in his speciality/rating he would circle A, "Fits very well."

Section 6 serves as an overall evaluation where the evaluator indicates whether he would recommend his selection for advisory training.

During the course of development, a decision was made that only Bureau of Naval Personnel detailers would use the information contained on the NAPR form and that the NAPR would not become a part of the individual's official record. Although the completed document could be used for a variety of purposes, this restriction was imposed for several reasons. If the reporting system is to function and if the assessments are to be true indications of an individual's probable success as an advisor, then the completed NAPRs should be used for advisor selection only. The data obtained from the evaluation instrument will not be used for other personnel management programs such as future assignment, promotion, or advanced training. The rationale behind this decision was based on past research in performance rating. This research had shown that a main source of error with most evaluation instruments is the tendency for the evaluator to assign a higher rating than is warranted. Most raters are reluctant to rate an individual low or even average on any scale which becomes a permanent part of his official record. As a result, personnel assessment evaluation reports

FOR OFFICIAL USE ONLY
(When completed)

NAVY ADVISOR PROFILE REPORT
NAVPERS 1300/8 (7-71)

REPORT BUPERS 1300-24

PART I - IDENTIFICATION DATA (To be completed by personnel officer)

1. NAME OF APPLICANT (Last, first, middle)	2. GRADE/RATE
3. SOCIAL SECURITY NUMBER	4. PRESENT SHIP OR STATION

PART II - (To be completed by rater)

1. HOW CLOSELY HAVE YOU OBSERVED THIS APPLICANT

☐ CLOSELY ☐ FREQUENTLY ☐ INFREQUENTLY

2. HOW LONG HAVE YOU KNOWN THIS MAN?

YEARS	MONTHS

3. HAVE YOU SERVED AS ADVISOR TO A FOREIGN COUNTRY?

☐ YES ☐ NO

COUNTRY	DATES OF SERVICE FROM	TO

4. FACTORS

(Read factor definitions in Instruction Manual)

	FACTORS	PREDICTED PERFORMANCE SCALE HIGH		MEDIUM			LOW	
1	PATIENCE, PERSISTENCE	A	B	C	D	E	F	G
2	TACT, DIPLOMACY, SOCIAL SKILL	A	B	C	D	E	F	G
3	ADAPTABILITY	A	B	C	D	E	F	G
4	SELF-RELIANCE, RESOURCEFULNESS, INGENUITY	A	B	C	D	E	F	G
5	FRIENDLINESS, SENSE OF HUMOR, SOCIABILITY	A	B	C	D	E	F	G
6	EMPATHY	A	B	C	D	E	F	G
7	MORALITY	A	B	C	D	E	F	G
8	EMOTIONAL STABILITY	A	B	C	D	E	F	G
9	INSTRUCTIONAL ABILITY	A	B	C	D	E	F	G
10	JOB DEDICATION, MOTIVATION	A	B	C	D	E	F	G
11	LEADERSHIP, ORGANIZATIONAL ABILITY	A	B	C	D	E	F	G
12	OVERALL EFFECTIVENESS AS AN ADVISOR	A	B	C	D	E	F	G

THE COMPLETED REPORT IS NOT TO BE SHOWN TO THE RATED INDIVIDUAL

Fig. 3.--Navy Advisor Profile Report

FOR DETAILING PURPOSES ONLY - WILL NOT BE INSERTED
INTO OFFICIAL RECORD OF INDIVIDUAL BEING RATED

FOR OFFICIAL USE ONLY
(When completed)

FOR OFFICIAL USE ONLY
(When completed)
NAVPERS 1300/8 (7-71)(BACK)

REPORT BUPERS 1300-24

5. GENERAL QUALITIES HOW WELL DOES EACH OF THE FOLLOWING WORDS OR PHRASES FIT THIS INDIVIDUAL?	OBSERVED PERFORMANCE SCALE						
	FITS VERY WELL		FITS FAIRLY WELL		DOESN'T FIT VERY WELL		DOESN'T FIT AT ALL
1 TECHNICALLY PROFICIENT IN HIS SPECIALITY/RATING	A	B	C	D	E	F	G
2 WILLING TO LISTEN AND LEARN	A	B	C	D	E	F	G
3 GOOD AT HANDLING PEOPLE	A	B	C	D	E	F	G
4 RACIALLY NON-PREJUDICED	A	B	C	D	E	F	G
5 JACK-OF-ALL-TRADES	A	B	C	D	E	F	G
6 MATURE IN JUDGMENT AND ACTIONS	A	B	C	D	E	F	G
7 ABLE TO TAKE CARE OF HIMSELF	A	B	C	D	E	F	G
8 USES COMMON SENSE	A	B	C	D	E	F	G
9 WELL ROUNDED NAVY KNOWLEDGE	A	B	C	D	E	F	G
10 HAS PRIDE IN APPEARANCE, ACTION AND ORGANIZATION (Professionalism)	A	B	C	D	E	F	G
11 PERFORMS WELL WITHOUT SUPERVISION	A	B	C	D	E	F	G
12 ABSENCE OF SUPERIOR ATTITUDE	A	B	C	D	E	F	G

6. RECOMMENDATION: If you had the authority and responsibility to do so, would your recommend his selection for advisory training? ☐ YES ☐ NO (Explain below)

7. COMMENTS: If desired, make specific comments regarding his strengths or weaknesses as a potential advisor.

8. TYPE OR PRINT NAME, GRADE/RATE OF RATER	SIGNATURE	DATE

FOR OFFICIAL USE ONLY
(When completed)

1--- PATIENCE, PERSISTENCE

The effective advisor in Vietnam needs a large reserve of patience to continue pursuing his assigned duties in the face of frequent and sometimes lengthy delays between the time he offers his advice and the time it is clearly accepted - or rejected. Without show of annoyance or anger, he must keep on offering and re-offering his suggestions until he is reasonably sure his counterpart has at least understood him -- and his counterpart's polite agreement to almost everything will give him few clues. Once understood, he must tolerate the possibility of having his ideas wholly ignored or rejected -- often without knowing why. If his advice is apparently accepted, he must be prepared for another waiting period until action is taken on it -- or even endure the disappointment of seeing no action at all.

More patience is required because of the advisor's position as 'the man in the middle' between the standards and organization of the U. S. Navy and those of the Vietnamese Navy. His American superior may issue the advisor one type of order, while his VNN counterpart has been given a conflicting order from his superior -- and both orders on the same subject. For the American, an action to be done on Tuesday should be done on Tuesday; for the Vietnamese, an action to be done on Tuesday will be done when the 'signs are right' for that action.

Patience and persistence are again needed after the advisor is confronted by delays and obstacles, because - if he is to be effective at all - he must then follow through to discover some of the reasons for his advice being rejected so that he can change his approach and try again.

* * * * *

On the Navy Advisor Profile Report, circle the letter on the Predicted Performance Scale which best describes how much patience and persistence you think this applicant would have as a naval advisor in Vietnam.

A--- Remains controlled, poised, and well-mannered in any and all situations. When confronted by setbacks and delays, he would continue his efforts without losing either his patience or his perspective.

B

C

D--- Shows some irritation or resentment at times. Would probably complain to his superior, or to fellow Americans, when things don't go his way. However, after some cooling off, he would continue his efforts and regain his composure and perspective.

E

F

G--- Insists that everything be done exactly how and when he says so. When others do not comply, he would lose his temper and would tell people just what he thinks. If he feels a job has to be done, he would probably do it himself so that it would be "right".

Fig. 4--Definition for the factor "Patience & Persistence"

have the traditional problem of being excessively inflated and, therefore, their value as evaluative instruments is less than desired. The evaluator would be less inclined to overrate an individual for RVN assignment if he were assured that his evaluation would not be used for any other purpose. Also the evaluator would be more honest in his ratings if the NAPR would not be shown to the rated individual.

This very briefly describes the main characteristics of the Navy Advisor Profile Report. Time does not permit going into the full developmental procedures. However, the documentation, the form, and instruction manual are contained in a research report which will be ready for distribution in October. Individuals interested in obtaining the report may write to the Naval Personnel Research and Development Laboratory, Bldg. 200, Washington Navy Yard, Washington, D. C. 20390.

REFERENCE

Yellen, Ted M. I., and McGanka, John F. The Navy Advisor Profile Report. Washington, D.C.: U. S. Naval Personnel Research and Development Laboratory. (In preparation).

FACTORIAL PROFILE OF THE E-8/E-9 EXAMINATIONS

ERLING A. DUKERSCHEIN

U.S. NAVAL EXAMINING CENTER
GREAT LAKES, ILLINOIS

New definitions of the role of senior and master chief petty officers led to the development of a new format for the subject matter area of certain E-8 and E-9 advancement examinations.

During and after this development, studies were started to evaluate its progress and the initial results (Macaluso, C. J., 1969), (Dow, A.N., Macaluso, C.J., 1970). These studies are a continuing effort and the subject of this paper represents the latest evaluation.

Ten ratings were selected for this study and the examination results are derived from the Series 57 (February 1971) examination cycle. Of the ten ratings selected, five represented the Old Format group and five the New Format group.

An inspection of the mean Navy Basic Battery scores of the two groups indicated no practical differences in basic ability, although the rates at Pay Grade E-9 had higher scores than those at E-8. The sample consisted of about 7900 candidates and comprised about 37% of all participating candidates.

The basic data for the study consisted of 20 section intercorrelation matrices representing 5 rates at E-8 and 5 rates at E-9 for the Old Format group; and 5 rates at E-8 and 5 rates at E-9 for the New Format group. The Old Format examinations at both E-8 and E-9 consist of six sections covering Occupational Knowledge, Military Qualification, Supervision and three aptitude areas; Mechanical Comprehension, Verbal Analogies and Arithmetic Reasoning. The New Format examinations at E-8 consist of six sections covering Occupation Knowledge, Military, Supervision and three general aptitude areas, Special Aptitudes, Communications and Problem Solving. The New Format examinations at E-9consists of five sections covering Occupational Knowledge, Military, Administration, Communications and Problem Solving.

The ratings representing the Old Format group were Aviation Machinist Mate, Data Processing Technician, Hospital Corpsman, Personnelman and Yeoman. The ratings representing the New Format group were Boilerman, Damage Controlman, Electronic Technician, Radioman and Storekeeper.

The twenty matrices of section intercorrelations were subjected to a centroid factor analysis followed by a varimax rotation. The results of this analysis are indicated in Tables I through Tables IV. These tables represent average factor loadings for three common factors and several independent factors.

An analysis of the twenty rates indicated two consistent common factors. The first of these is consistently dominated by the professional section and it may be labeled "Specialized Knowledge". The military section tended to have significant loadings on this factor. The second common factor was dominated by the aptitude sections and may be labeled "General Abilities". Sections in the Old Format examinations that load heavily on this factor are Mechnical Comprehension, Verbal Analogies and Arithmetic Reasoning. In the New Format examinations the Special Aptitudes, Communications and Problem Solving sections have significant and consistent loadings on this factor.

In seventeen of the twenty analyses a third common factor occured. However, this factor was not defined by any consistent pattern from rate to rate. This factor on the average accounted for about twenty percent of the common variance. Although it cannot be considered to have a questionable existence, when it does occur it appears to be peculiar to the rate and pay grade.

In the area of the specific variance of sections several consistent trends occur over the rates and pay grades. The first trend of note concerns the professional section. In fifteen of the twenty rates analyzed the professional section showed significant specific or independent variance. This analysis did not consider the relationship between items within a section. Consequently, it is extremely difficult to provide a label for the specific variance of any particular section. However, if one notes the relatively low loadings of the aptitude sections on the common factor previously designated "Specialized Knowledge" one might conjecture that this specific variance is produced by items that require both specialized knowledge and the ability to apply that knowledge, neither requirement by itself being sufficient. If a future analysis indicates that this is the case this factor could conceivably be labeled a proficiency factor (Dukerschein, E.A., 1969).

The second trend concerns the Old Format group. In the ten rates comprising this group at both pay grades, the arithmetic reasoning section consistently displayed significant specific variance. In view of the previous work of Thurston and many others in this area over the years it is probably safe to label this variance as due to a numerical or number factor.

In the New Format group at pay grade E-8 three sections in addition to the professional section showed a significant amount of independent variance. These sections were the Special Aptitudes section, the Communications Section and the Problem Solving section. At pay grade E-9 the additional sections showing significant specific variance were the Communications Section and the Problem Solving section.

The design of the communications section was based on the work of William V. Haney (Haney, W. V., 1967). Broadly speaking, the communications section attempts to test whether a candidate realized when he is making an inference. The point being, if he does not realize when he is making an inference, he will not evaluate it in terms of its' probable accurance. This seems an important point in receiving or producing a written or an oral communication. Can you differentiate between what the communication actually says and what may be inferred. For want of a better phrase this specific variance might be termed an awareness of the difference between statements of observation and statements of inference.

In the design of the special aptitudes and the problem solving sections a broad spectrum approach was used. The items involve the usual verbal, quantitative and pictorial content. They require processes of cognition, intermediate production and evaluation; and the nature of the problems and their solutions involve classes and relationships, recognition of structured systems, transformation of given information and logical implications. Consequently, the specific variance of these sections undoubtedly represent complex factors which may be labeled combined abilities.

Let me close with a quotation from a previous study (Macaluso, C.J., Dow, A.N. 1969). "The Navy was looking for candidates who had retained their mental alertness, had not forgotten the details of their respective specialties, and who were still able to tackle problems, and could handle pragmatically problems in management and human relations."

It would appear that the New Format examinations support this quest.

TABLE I

Average Factor Loadings for Five Rates of Old Format Group at Pay Grade E-8

SECTIONS	COMMON FACTORS I	II	III	INDEPENDENT FACTORS S_1	S_2	S_3	S_4	S_5	S_6
1. Professional	.452*	.085	.131	.428*					
2. Military	.407*	.183	.086		.071				
3. Supervision	.219	.173	.302*			.130			
4. Mech. Comprehension	.085	.519*	.202				.268		
5. Verbal Analogy	.238	.557*	.146					.295	
6. Arithmetic Reasoning	.223	.547*	.171						.517*

TABLE II

Average Factor Loadings for Five Rates of New Format Group at Pay Grade E-8

SECTIONS	COMMON FACTORS I	II	III	INDEPENDENT FACTORS S_1	S_2	S_3	S_4	S_5	S_6
1. Professional	.417*	.201	.098	.397*					
2. Military	.084	.066	.265		.032				
3. Special Aptitudes	.259	.568*	.222			.540*			
4. Supervision	.190	.142	.199				.105		
5. Communications	.179	.385*	.323*					.575*	
6. Problem Solving	.228	.575*	.225						.378*

TABLE III

Average Factor Loadings for Five Rates of Old Format Group at Pay Grade E-9

SECTIONS	COMMON FACTORS I	II	III	INDEPENDENT FACTORS S_1	S_2	S_3	S_4	S_5	S_6
1. Professional	.508*	.157	.100	.395*					
2. Military	.370*	.035	.156		.219				
3. Supervision	.255	.047	.119			.138			
4. Mech. Comprehension	.146	.564*	.174				.292		
5. Verbal Analogy	.177	.356*	.333*					.285	
6. Arithmetic Reasoning	.117	.545*	.137						.545*

TABLE IV

Average Factor Loadings for Five Rates of New Format Group at Pay Grade E-9

SECTIONS	COMMON FACTORS I	II	III	INDEPENDENT FACTORS S_1	S_2	S_3	S_4	S_5	S_6
1. Professional	.497*	.092	.182	.348*					
2. Military	.294	.120	.316*		.249				
3. Administration	.139	.062	.306*			.032			
4. Communications	.163	.506*	.116				.622*		
5. Problem Solving	.095	.485*	.188						.612*

* = Factor loading of .300 or higher

REFERENCES

Dow, A.N., Macaluso, C.J. The new E-8 and E-9 exams. A first year report. In Proceedings of the 12th Annual Conference, Military Testing Association. 1970.

Dukerschein, E.A., Winiewicz, C.S. A classification system for achievement items based on candidate subgroup definition. In Proceedings of the 11th Annual Conference, Military Testing Association. 1969.

Haney, W.V. Communication and organizational behavior. Richard D. Irwin, Inc., Homewood, Ill. 1967.

Macaluso, C.J., Dow, A.N. E-8 and E-9 test: A new approach. In Proceedings of the 11th Annual Conference, Military Testing Association. 1969.

DEVELOPMENT OF A UNIVERSAL EQUATION FOR PREDICTING JOB DIFFICULTY

By
Donald F. Mead, Major, USAF
Headquarters Air Training Command
Randolph AFB, Texas

Introduction

Each of the military services is currently conducting occupational surveys which provide individual job descriptions and identify job types within each career ladder. A recurring need has been the development of a technique to derive an index of the relative difficulty of the jobs identified. A job difficulty index could be used (1) to develop increasing difficulty and responsibility as personnel progress in their career ladders; (2) to assist in establishing minimum aptitude requirements for positions and classes of positions; (3) to compare the difficulty level of work assigned to individuals at various aptitude levels; (4) to compare the difficulty level of work assigned to technical school graduates, individuals bypassing the technical school, and individuals receiving directed duty assignments; (5) to investigate the interaction between job difficulty, job satisfaction, and felt utilization of talent; (6) to determine the appropriate grade requirements for positions; and (7) to guide decisions about modifications in the classification structure.

Historically job difficulty has been associated with the values obtained with various job evaluation systems. A review of the major systems used in industry revealed that none lent themselves to the military situation with its vast number of jobs and wide dispersal throughout the world. The problem was to develop a new job evaluation approach, one which was quantitative, easy to administer, objective, and provided maximum interface with existing military occupational analysis data processing programs. A hypothesized approach was to have experienced supervisors arrange jobs in their career ladder according to their relative difficulty, and then apply Christal's (1967) policy capturing model to capture the judgment policy of the supervisor raters. The resulting prediction equation could then be applied to all jobs in the career ladder to derive their difficulty scores.

This presentation reviews the basic research design and results of three independent repetitive studies confirming the policy capturing approach and the development of a universal equation for establishing the difficulty level of jobs across all AF career ladders.

Method

The initial studies conducted to test this new job difficulty evaluation technique employed jobs from the Medical Materiel, Accounting and Finance, and Vehicle Maintenance career ladders. In each study the following research design was employed.

Selection and Arrangement of Job Descriptions

Two hundred fifty job descriptions, listing all tasks performed and the relative time spent performing them, were randomly selected in each study to serve as the criterion sample. These descriptions were placed in 16 separate random order listings, each containing all 250 criterion job descriptions. Each listing was then divided into 10 subsets of 25 job descriptions, for a total of 160 subsets. With this design each job description appeared in 16 subsets, each time with varied accompanying descriptions. The three career ladders studied were selected due to the dissimilarity in the nature and number of tasks performed in the jobs within each ladder.

Development of Job Difficulty Criterion Values

One hundred sixty, 7-skill level noncommissioned officers (NCOs) assigned to the career ladder studied were randomly selected from the Uniform Airman Record File to evaluate the job descriptions. Each NCO received a set of 25 job descriptions listing tasks performed and time spent performing each. Each NCO ranked the jobs according to their relative difficulty. A job difficulty criterion value was established for each job by computing its mean difficulty rank order value.

Development of Task Difficulty Values

For purposes of these studies task difficulty was defined as the time required to learn to perform the task satisfactorily. Occupational inventory booklets listing all tasks performed in each ladder studied were mailed to 7- and 9-skill level NCOs working in each ladder. The raters assigned a difficulty value to each task using a seven point relative scale. The task difficulty values were determined by computing the mean rating assigned by each group.

Development of the Predictor Variables

Prior to analysis, individual job descriptions were examined to identify variables which could have influenced the raters who provided the job difficulty criterion values. Eighteen hypothesized predictors, including linear, squared, and interaction variables, were selected for analysis.

TABLE 1

Definition of Predictor Variables

Variable Number	Variable Description
1	Mean Job Difficulty Rank Value (Criterion). Mean rank order position computed for each of the 250 criterion job descriptions using ranks assigned by NCOs working in the Career Ladder.
2	Number of Tasks Performed. The number of tasks listed as being performed on each criterion job description.
3	Mean Task Difficulty, 9-Level Ratings. This variable used mean difficulty values for each task obtained from 9-skill level NCOs using a 7-point relative scale. This value was computed for each job description by summing the task difficulty means for the tasks performed and dividing by the number of tasks in the job description.
4	Mean Task Difficulty, 7-Level Ratings. Same as Variable 3 except mean task difficulty values from 7-skill level NCOs were used.
5	Variable 3 plus Variable 4. Same computation as used in Variables 3 and 4 except the task difficulty values reflect the combined judgments of 7- and 9-skill level raters.
6	Average Difficulty of Tasks Performed per Unit Time Spent, 9-Skill Level. Computed for each job description by summing the cross products of 9-skill level mean task difficulty values by time spent on the tasks performed.
7	Average Difficulty of Tasks Performed per Unit Time Spent, 7-Skill Level. Identical to Variable 6 except 7-skill level mean task difficulty values were used.
8	Variable 6 plus Variable 7. Identical computation as found in Variables 6 and 7 except the mean task difficulty values reflect the combined judgments of 7- and 9-skill level raters.
9	Job Difficulty-Average Grade. This predictor was generated by summing the average grade level task values for the tasks performed in each job description and dividing by the number of tasks performed.
10	Range of Task Difficulty. A generated variable obtained by computing the standard deviation for each job using the mean task difficulty values obtained from 9-skill level raters.
11	Variable 2, Squared
12	Variable 5, Squared
13	Variable 8, Squared
14	Variable 9, Squared
15	Variable 10, Squared
16	Variable 2 times Variable 8
17	Variable 2 times Variable 10
18	Variable 8 times Variable 10
19	Variable 2 times Variable 8 times Variable 10

Capturing Supervisors' Judgment Policy

To capture the supervisors' job difficulty judgment policy, the predictor variables were analyzed through a series of multiple regression problems. The predictors were grouped logically, and the R and R^2 values were determined. The effect of adding or subtracting variables from these groupings was evaluated by noting the change in criterion variance (R^2) accounted for by such modifications. These computations revealed the predictors and their associated weights which most accurately reproduced the supervisors' job difficulty evaluations.

Results

Job Difficulty Criterion Values

In each of the studies job difficulty criterion measures were derived by computing the mean of the rank order positions assigned each job by supervisors in the ladder examined. The number of ranking judgments per job description varied from 8 to 16, with most receiving 10 or more. The estimated reliability of these judgments was computed using Lindquists' (1953, p361) intraclass correlation technique. The Spearman-Brown formula was applied to the derived values to obtain reliability estimates for various sample sizes. As seen in Table 2, the results in each study indicate the derived criterion values are stable, reliable measures and that there is a definite evaluation policy to be captured.

TABLE 2

Estimated Reliability Coefficients for Mean Job Difficulty Ranking by Various Rater Samples

	Estimated Reliability (r_{kk})		
Rater N	Accounting and Finance	Vehicle Maintenance	Medical Materiel
r_{11}[a]	.52	.43	.60
6	.87	.82	.89
7	.88	.84	.91
8	.90	.86	.92
9	.91	.87	.93
10	.92	.88	.93

[a]Mean reliability for one rater (sample value).

Task Difficulty Ratings

Task difficulty values for each identified task in each career ladder were derived by computing the mean scale value assigned by the 7- and

9-skill level raters. Estimates of reliability for each rating group and for the combined groups were computed using the Lindquist intraclass correlation technique.

TABLE 3

Estimated Reliability of Task Difficulty Ratings

Rating Group	Accounting and Finance		Vehicle Maintenance		Medical Materiel	
	N	r_{kk}	N	r_{kk}	N	r_{kk}
7-Skill Level	15	.85	22	.86	10	.91
9-Skill Level	20	.89	17	.91	10	.93
Total	35	.93	39	.94	20	.96

As shown in Table 3, the mean task difficulty values used in the three studies are stable, reliable measures. Further, these results indicate that both 7- and 9-skill level NCOs provide stable measures of task difficulty using the "time required to learn" definition of task difficulty and the 7-point difficulty scale employed in these studies.

Development of Job Difficulty Prediction Equation

In each of the three development studies the predictor variables were analyzed through a series of multiple regression problems. This identified the most efficient combination of variables which best predicted the job difficulty criterion values. In each study the derived policy equation contained the same three predictor variables. These were (1) number of tasks performed, (2) task difficulty per unit time, 7-skill level ratings, and (3) number of tasks performed, squared. The correlation between the predicted job difficulty values and the criterion values for each career ladder studied is shown in Table 4.

TABLE 4

Correlation Between Predicted and Criterion Job Difficulty Values

Ladder	R	R^2
Accounting and Finance	.95	.90
Vehicle Maintenance	.93	.86
Medical Materiel	.95	.90

Stability of Derived Equation

In each development study the stability of the derived policy equation was tested. The job descriptions were divided into two equal development samples using the odd-even technique. Least square weights were computed separately for the three predictor variables within each group. Each evolved equation was used to predict the difficulty level of jobs within its own development sample and also cross-applied to predict the difficulty level of jobs in the opposite sample. In each study non-significant differences were found between correlations obtained in the cross-application and total sample analysis.

To test the total sample three-variable R for the possibility of inflation resulting from chance errors, Garretts' (1958, p416) correction formula for shrinkage was applied in each investigation. In each instance a non-significant correction was found. Since the correction was non-significant for both the shrinkage and cross-application tests, it appears that the job difficulty values obtained in each study with the three-variable prediction equation were stable measures. Analyzing the predictor variables emerging in each policy equation indicates that three factors apparently accounted for the supervisors' judgments: (1) the number of tasks appearing in the job description, (2) the difficulty of the tasks performed, and (3) the time spent performing the tasks.

Development of a Universal Equation

Although the independent development studies involved diverse job activities, the same three predictor variables combined to duplicate successfully the supervisors' job difficulty judgments. The similarity of the three developed policy equations suggested the possibility of developing a universal or constant standard weight equation to predict the difficulty level of jobs across Air Force career ladders. To derive job difficulty values, relevant predictor variable information could be collected concurrently with occupational analysis survey data. Applying the constant standard weights to these predictor data, the appropriate raw score regression weights would be obtained to derive the predicted job difficulty indices. Such a technique would reduce the cost and time presently spent in securing this information.

This concept was tested by employing information from the three completed development studies. In each investigation standard score beta weights were developed for the three variables to maximize the correlation between the predicted values and supervisors' evaluations (criterion measures). As shown in Table 5, the standard score weights for the identified variables were quite uniform.

TABLE 5

Standard Score Weights for Selected Predictor Variables

Career Ladder	Standard Score Weight: Variable 2	Variable 7	Variable 11	Criterion Standard Deviation
Medical Materiel	1.12582776	.45263499	—.58673349	5.7705
Vehicle Maintenance	1.29125838	.51612430	—.61529753	4.9992
Accounting and Finance	1.58510913	.39230372	—.95835786	5.4198
Mean	1.33406509	.45368767	—.72012963	5.3965

Note. -- Variable 2: Number of Tasks Performed
Variable 7: Task Difficulty per Unit Time Spent, 7-Level Ratings
Variable 11: Number of Tasks Performed, Squared

The similarity of these values suggested that the most suitable constant standard score values would be derived by computing the mean beta weight for each predictor. These constant standard score weights were then applied to the standard deviations of the predictors in each study to obtain raw score regression weights. The formula for this conversion was:

$$b_X \text{ regression weight} = \overline{(SSW)} \left(\frac{\overline{SD} \text{ Criterion}}{X \text{ Standard Deviation}} \right)$$

where

b_X = computed variable raw score regression weight for the career ladder studied

$\overline{SSW}$ = mean standard score weight from Table 5

X SD = standard deviation of the predictor obtained in the career ladder analyzed

$\overline{SD}$ = mean criterion standard deviation value from Table 5

This conversion was made for each of the three predictor variables in each development study. A constant mean criterion standard deviation value was employed since criterion measures would not be available in the normal application of this technique. This approach was validated by applying the raw score regression weights to the 250 jobs in each development study to derive new predicted difficulty indices which were correlated with the

respective criteria. The correlation between these derived difficulty values and those obtained using the original development sample beta weights are shown in Table 6.

TABLE 6

Comparative Efficiency of Prediction Equations Using Development Sample Beta Weights and Constant Standard Weights

(Criterion: Supervisor Ratings of Job Difficulty)

	R for Predicted Job Difficulty		
Career Ladder	Development Sample Beta Weight Equation[a]	Constant Standard Weight Equation[a]	Significance of Difference[b]
Medical Materiel	.9486	.9479	.110
Vehicle Maintenance	.9269	.9247	.155
Accounting and Finance	.9511	.9460	.353

[a]For 250 jobs in each career ladder.

[b]1.96 needed for significance at the .05 level.

These results indicate that valid job difficulty values may be obtained with the derived universal equation. In each instance the correlation computed from the constant mean standard weights did not differ significantly from the one obtained with the development sample beta weights. In effect, the universal equation has reproduced values which originally required the assistance of approximately 400, 7- and 9-skill level Air Force NCOs. Information concerning the number of tasks performed is obtained during an occupational survey. The only predictor information missing is the task difficulty evaluations, which have been reliably obtained from as few as 20 NCOs assigned to the ladder analyzed.

This new technique and the universal equation has been cross-validated with jobs in ladders not involved in their initial derivation with the following results.

TABLE 7

Efficiency of the Universal Job Difficulty Equation

Specialty	Name of Specialty	Least Squares R	Universal Equation R
811XX	Security Police	.922	.914
702XX	Administrative	.977	.970
647XX	Materiel Facilities	.942	.930
645XX	Inventory Management	.936	.917
631XX	Fuel Services	.942	.938
605XX	Air Transportation	.930	.925
571XX	Fire Protection	.939	.888
551XX	Civil Engineering, Pavements	.928	.925
543XX	Electrical Power Production	.937	.923

Summary and Conclusions

In three independent development studies, the job difficulty evaluation policy of supervisors has been captured and their judgment decisions simulated with multiple regression equations. In these investigations the correlations between predicted job difficulty values and supervisors' judgments were .95 for Medical Materiel jobs, .93 for Vehicle Maintenance jobs, and .95 for Accounting and Finance jobs.

In each of these studies the same three predictor variables combined in the multiple regression equation to capture the supervisors' evaluation policy. In making their job difficulty decisions, the supervisors apparently considered the number of tasks in the job description, the time required to learn the tasks performed, and the amount of time the incumbents spent performing each task.

Reliable measures of task difficulty were obtained by using a 7-point relative scale and defining difficulty as time required to learn to perform the task satisfactorily. Reliable job difficulty criterion measures were obtained for each ladder by computing the mean rank order position assigned jobs by experienced supervisors. In each instance reliability estimated in excess of .90 were obtained with 10 or more rankings per job.

In each investigation a three-variable regression equation was developed which successfully simulated the supervisors' job difficulty evaluations. In each study the correlation between predicted values and criterion measures was R=.93 or higher. A negligible correction was found when the correlations were tested for shrinkage. These studies indicated that valid job difficulty measures could be obtained through the policy-capturing approach.

In each of the completed studies, the supervisors' evaluation policy was converted into a standard score regression equation which maximized the correlation between predicted and criterion job difficulty values. These three equations were converted into a proposed constant standard score or universal equation representing a composite supervisor job evaluation policy. Validity of the equation was tested by applying the weights to predictive data from each development study to derive appropriate raw score regression weights. New predicted job difficulty values were computed for the jobs in each ladder and correlated with their respective criteria. Non-significant differences in correlations were found between the criterion measures and simulated difficulty values obtained with development sample beta weights and the universal equation beta weights. The results indicated that the universal equation accurately captured the supervisors' evaluation policy and yielded valid job difficulty scores. The universal equation has been cross validated in several additional career ladders with similar success. It appears that jobs across Air Force career ladders can be evaluated with the universal equation. The cost and man-hour savings are readily apparent. Perhaps the most outstanding merit of the system is its quantitative format which lends itself to advanced computerized personnel management models.

REFERENCES

Christal, R. E. Selecting a harem - and other applications of the policy-capturing model. PRL-TR-67-1, AD-658-025. Lackland AFB, Texas: Personnel Research Laboratory, Aerospace Medical Division, March 1967

Garrett, H. E. Statistics in psychology and education. New York: David McKay Co., Inc., 1958.

Lindquist, E. F. Design and analysis of experiments in psychology and education. Boston: Houghton Mifflin Co., 1953.

Mead, D. F. Development of an equation for evaluating job difficulty. AFHRL-TR-70-42. Lackland AFB, Texas: Personnel Division, Air Force Human Resources Laboratory, November 1970.

Mead, D. F. Continuation study on development of a method for evaluating job difficulty. AFHRL-TR-70-43. Lackland AFB, Texas: Personnel Division, Air Force Human Resources Laboratory, November 1970.

Mead, D. F. and Christal, R. E. Development of a constant standard weight equation for evaluating job difficulty. AFHRL-TR-70-44. Lackland AFB, Texas: Personnel Division, Air Force Human Resources Laboratory, November 1970.

TASK DIFFICULTY AND TASK APTITUDE BENCHMARK SCALES AN EXPLORATORY STUDY

By

Squadron Leader John W. K. Fugill, USAF (RAAF)
Occupational and Career Development Branch
Personnel Research Division (AFHRL)
Lackland Air Force Base, Texas

I. Introduction

This paper is a report on an attempt to develop experimental Task Difficulty Benchmark Scales (TDBS) and Task Aptitude Benchmark Scales (TABS).

For at least twenty years, the entry of enlisted personnel into USAF career ladders has been determined by aptitude requirements established originally on the basis of judgement. As a function of recruiting and training objectives, an Aptitude Index predicts the probability of success in training on an actuarial basis (Harding & Brokaw, 1958). As successively higher scores are achieved on a given aptitude index, the probability of the individual's success in training rises. Thus, the raising of a minimum level from 60 to 80 may reduce the failure rate to zero, but may establish an unacceptably low selection ratio for the recruiting organization. On the other hand, reduction of an aptitude score requirement from 80 to 60 would double the number of individuals eligible for a particular specialty.

Although this approach can be defended on purely functional grounds, there is no substantial body of theory to justify its continuance. In the post World War II period, the USAF has had no difficulty in recruiting an adequate share of highly talented airmen; therefore, the need for a more refined method has not been urgent. However, in a zero-draft environment, the manpower resources required to fill high-aptitude enlistment quotas may be more limited than is the case now (Vitola & Valentine, 1971). What is needed for the future is a systematic method for the objective determination of aptitude levels as functions of task difficulty. To this end, TDBS and TABS have been the subject of experimentation at Personnel Research Division, Air Force Human Resources Laboratory.

Insofar as judgments about difficulty and aptitude can be anchored to a common frame of reference, benchmark scales appear to have several practical applications.

1. Determination of the relative difficulty of all tasks across <u>all</u> career ladders.

2. Computation of realistic aptitude cutting scores, and ultimately, the re-alignment of aptitude scales.

3. Identification of career areas in which a reduction in the aptitude requirement would least jeopardize mission effectiveness.

4. Computation of the percentage of a defined base population capable of performing a given task.

Obviously these applications are related to decision-making for the selection and training of first-term airmen. For example, contingency plans could be developed to counteract a decline in the recruitment of high-aptitude applicants.

II. Purpose of the Study

This exploratory study has a sixfold purpose.

1. Clarification of the concepts of task difficulty and task aptitude.

2. Determination of the reliability of work supervisors' judgments about the relative difficulty of selected tasks in the "mechanical" field.

3. Determination of the reliability of Behavioral Scientists' judgments about the relative aptitude required to learn to perform those same tasks.

4. Construction of an experimental TDBS which could be used to determine the relative difficulty of all tasks performed in the "mechanical" career ladders.

5. Construction of an experimental TABS which could be used to determine the relative aptitude required to learn to perform those same tasks satisfactorily.

6. Determination of any significant relationship between the TDBS and the TABS.

III. Literature on Task Difficulty

For the most part, task difficulty has been defined operationally in terms of specific learning situations. For example, in target tracking, difficulty was defined in terms of target size (Barch, 1953);

in electronic fault-finding, the essential difficulty was the choice of the best strategy (Dale, 1957); for perceptual motor skills, difficulty was defined as the number of response alternatives (Deupree & Simon, 1963). In some studies, the words "difficulty" and "complexity" are used synonymously and inappropriately; in other studies, difficulty is said to be a function of complexity.

Day (1956) summarized the principal findings of experiments concerned with the effect of task difficulty on the transfer of training. It was observed that definitions of difficulty were inconsistent, and that the sources of difficulty were inadequately controlled. Holding (1962) concluded that difficulty was not a useful category for the prediction of transfer efficiency. Tilley (1969) stated that information-processing models indicate that the concept of task difficulty is multi-dimensional; that tasks should be described by a profile of characteristics rather than by some global index of difficulty; and that training can be effective only if the source of difficulty of a task is identified correctly.

Shaw (1963) describes some procedures used in collecting group tasks, identifying task dimensions, and scaling tasks along those dimensions. The dimension of difficulty was defined as the amount of effort required to complete a task; it was hypothesized that difficulty is influenced (or perhaps determined) by the number of operations, skills, and knowledges required for successful task completion. Shaw concluded that the difficulty dimension was relatively stable and strong in the sense that judgments were consistent and the factor structure was relatively stable. In regard to scaling procedures, it was observed that judges showed greater agreement on tasks near the extremes of the dimension than on those near the middle categories; and that consistent scale values could be obtained with few judges.

Lecznar (1971) evaluated three methods of estimating task difficulty in the USAF Medical Materiel Career Ladder (915X0). Because of the problem of providing judges with the same structured concept of task difficulty, difficulty was defined simply as "time needed to learn to do the task satisfactorily." One conclusion relevant to this study is that considerable agreement can be achieved among both supervisor raters and supervisor rankers as to the relative difficulty of tasks, with factors such as training, aptitude and experience held constant.

The studies cited above are only obliquely relevant to benchmark scale construction. That the research on task difficulty has been of a fragmentary nature helps to explain the general deficiency of theoretical formulations.

IV. Conceptual Problems

An underlying hypothesis is that the concept of aptitude requirement is essentially a function of task difficulty. This one sentence embodies the main conceptual problem of this study: i.e., the slipperiness of the concepts of task aptitude and of task difficulty, but particularly of the latter. Definitions of aptitude show agreement on three points:

1. an innate ability factor,

2. ability derived from accumulated experience, and

3. some characteristics symptomatic of ability to learn specified knowledge or skills (Warren, 1934).

The acceptance of the last-mentioned point is essential to the acceptance of the definition of difficulty which is to follow.

Initially, task difficulty was considered simply in terms of difficulty in performing a task satisfactorily under conditions defined as normal. However, discussions with experienced supervisors appear to support Madden's view (1962) that a worker seldom perceives difficulty as a quality of a task (job) once it has been learned. Difficulty is attributed to such things as poor conditions of work, lack of practical experience, and interpersonal frictions. Clearly, the difficulty of a given task is not perceived as a constant. A job description printout may show that 45 workers in a hundred perform a particular task. But, there are, in fact, 45 tasks of varying perceived levels of difficulty with no necessarily common factor to account for the difficulty. All things considered, the concept of 'task performance difficulty' is so nebulous that no satisfactory definition can be formulated now.

A second definition of difficulty to be considered is 'time needed to learn to do a task satisfactorily' (Lecznar, 1971). Again, difficulty is not perceived as a constant, but as a function of personal and environmental factors, e.g., ability deficiencies and unsatisfactory instruction (Madden, 1962). Some tasks will be learned in formal classroom training, some through on-the-job training, others through self-instruction, and a few through undirected experience. (This comment is a necessary qualification of the third element of the definition of aptitude (Bingham, 1937). It is obvious that the assignment of difficulty values across such differing learning backgrounds requires fine judgment. Studies by Mead (1970a, 1970b), Mead and Christal (1970) and Lecznar (1971) indicate that work supervisors achieve a high level of agreement about relative task difficulty when guided by this definition.

This second definition (time to learn) was chosen after a consideration of the methodological problem: what policy instructions should be given to the judges? The two alternatives have widely differing implications. First, "difficulty in performance" :- the rankers' judgments might be influenced by their knowledge and experience of such things as physical working conditions and interpersonal relations. Second, "time taken to learn" :- the rankers' judgments might be influenced by their knowledge and experience of such things as training methods and individual abilities. A third alternative - no policy instructions - may well have resulted in high interrater agreement, but would have provided no information for the clarification of the concept of difficulty. To summarize: insofar as 'ability to learn' is a concomitant of aptitude, and difficulty and aptitude appear to be conceptually complementary, the definition 'time to learn' appears to be the least objectionable.

V. Method

Within the USAF Airmen Classification Structure, four work groups were defined at the journeyman level to coincide with established aptitude indexes: mechanical, electronic, administrative, and general. Thus the mechanical work group contains tasks from career ladders for which mechanical selector aptitude indexes are prescribed (AFM 35-1). Selected supervisors in this broad mechanical area wrote task statements to the following specifications:

1. Level of Skill: Because the ultimate aim is the re-alignment of aptitude scales for training at the point of entry into the service, the tasks were limited to those performed at the 3 and 5 levels of skill. No supervisory tasks were included.

2. Specificity: Most statements used in occupational analysis studies were considered to be too general insofar as a single action word covered a number of related or subsidiary tasks. For example, "tune automobile engine" is a general statement; "dress spark plug electrodes with a file to secure flat parallel surfaces on both electrodes" is a subsidiary task described specifically. Writers were instructed to use specific action words; to name equipment and tools; and to indicate the size or scope of the activity, permissible tolerances, degree of precision required, and any restrictive factors which obviously contributed to difficulty.

3. Comprehensibility: The agreed upon criteria were: use of simple words as opposed to unnecessarily difficult words; correct use of technical terms; and clarity and conciseness of expression.

4. Range of Difficulty: A wide range of difficulty was a crucial requirement. The higher level-of-difficulty items presented special problems with regard to specificity and comprehensibility.

Each selected task statement was printed separately on a card; the cards were then ordered randomly into sets of 165. These sets were mailed to two groups of mechanical work superintendents (9-skill level) in the field. Given the operational definition of difficulty as "time needed to learn to do a task satisfactorily," Group A rated the statements on a 7-point scale from 1 (very much below average time) to 7 (very much above average time). Group B, with the same policy instructions, rank-ordered the statements from 1 (easiest) to 165 (most difficult) with no ties permitted. Forty-four sets of cards were returned by the superintendents; 20 sets were rank-ordered and 24 sets were rated. In addition, twelve behavioral scientists experienced in occupational and career development research rank-ordered the same 165 statements on the degree of aptitude required to learn to do a task satisfactorily.

To the three sets of raw data several computerized statistical procedures were applied.

VI. Preliminary Results and Discussion

Inter-rater Reliabilities: using the intraclass correlation technique described by Lindquist (1953), estimates of reliability were computed. To obtain estimates for various sample sizes, the Spearman-Brown formula was applied to the single judges' reliability coefficients. These are shown in Table 1.

Table 1. Estimated Reliabilities

	N	r_{kk}
Difficulty Rankings	r_{11}	.592
	6	.897
	8	.921
	10	.936
	12	.946
	20	.967
Difficulty Ratings	r_{11}	.485
	6	.850
	8	.883
	10	.904
	12	.919
	24	.958
Aptitude Rankings	r_{11}	.684
	6	.929
	8	.945
	10	.956
	12	.963

These indicate that very high agreement can be achieved by superintendents and behavioral scientists with the dimensions of difficulty and aptitude respectively.

Correlation Coefficients: rank-difference correlation coefficients were computed between the pairs of scales based on the rank-ordering by means and the rank-ordering by variance (SD). These are shown in Table 2.

Table 2. Correlation Coefficients

	Difficulty Ranking Difficulty Rating	Difficulty Ranking Aptitude Ranking	Difficulty Rating Aptitude Ranking
Rank-order by means	0.907	0.891	0.884
Rank-order by variance	0.413	0.361	0.338

The correlation coefficients of the rank-ordered means are sufficiently high to confirm the common-sense notion that an aptitude

requirement is a function of task difficulty. Also, they indicate a marked positive relationship between the task positions on the TDBS and the positions on the TABS.

Variance: also from Table 2, the coefficients of variance indicate that judges of task difficulty and judges of task aptitude have moderate agreement about their disagreements. The full significance of this has yet to be determined. In keeping with Shaw's findings, it is clear that there is less variance at the ends of the scales, but notably less at the lower (easy) end. The following examples show varying degrees of variance across the three scales.

Example 1.

Refinish rubbing surfaces of disc brakes (automobile) using precision equipment.

	Difficulty Rating	Difficulty Ranking	Aptitude Ranking
Position	97/165	94/165	93/165
SD	(0.93)	23.27	19.65

This item fell into central positions on all three scales, and the variance was relatively low.

Example 2.

Test two-way radio to ensure proper transmission and reception.

	Difficulty Rating	Difficulty Ranking	Aptitude Ranking
Position	7/165	24/165	52/165
SD	(1.13)	38.27	36.96

This item, intended to represent a simple task relevant to both the mechanical and the electronic work fields, shows a lack of agreement between and within the three groups of judges. One superintendent's comments underscore the statement's ambiguity: "What sort of radio? Does the task include FCC Class III restrictions? Does it only require depressing a microphone switch and talking, or does it require changing of frequencies or tuning?" This illustrates the need for greater rigor in the writing of statements if the complementary requirements of specificity and comprehensibility are to be met.

Example 3.

Write a 500 word memorandum stating a factual work problem and giving ideas to solve the problem (from journeyman to immediate supervisor).

	Difficulty Rating	Difficulty Ranking	Aptitude Ranking
Position	160/165	146/165	165/165
SD	(1.25)	45.44	4.75

This experimental item is not strictly "mechanical," but could be performed in mechanical career fields. It could be used also for the electronic, administrative and general scales. Behavioral scientists were in strong agreement that it required the highest aptitude; for the superintendents who rank-ordered the tasks, it had the highest variance on their scale of difficulty.

At this point, two primary sources of variance may be hypothesized.

1. Personal factors - such as the extent of the judges' knowledge of the tasks, and their perceptions of the social values of those tasks.

2. Task statement factors - such as specificity and comprehensibility.

Distribution of Difficulty: the frequency distribution of the 165 tasks across the 7 point difficulty rating scale is as follows:

4	24	44	47	37	9	0
0.5-1.49	1.5-2.49	2.5-3.49	3.5-4.49	4.5-5.49	5.5-6.49	6.5-7.49

The skewness indicates that there are more tasks with the mean difficulty at the lower end and center of the scale than at the top of the scale. This suggests that there was a lack of agreement on the difficult tasks and/or a higher representation of easier tasks in the sample. In preparation for a replication of this study, some of the tasks contributing to the lumpiness of the distribution will be removed to achieve approximate rectilinearity.

VII. Conclusions

For this exploratory study, two major hypotheses were proposed.

1. The concept of aptitude requirement is essentially a function of task difficulty.

2. Judgments about task difficulty and task aptitude can be anchored to a common frame of reference.

Using the operational definition of difficulty as 'time taken to learn to do a task satisfactorily', there is *prima facie* evidence to support both hypotheses as far as the mechanical work field is concerned. Once the inter-related problems of specificity, comprehensibility, and distribution of difficulty have been reduced to manageable limits, highly reliable TDBSs and TABSs may be constructed with comparative ease. With these instruments occupational researchers may, with high confidence, make objective judgments about the relative aptitudes required to learn to perform a variety of mechanical tasks.

REFERENCES

Barch, A. M. The effect of difficulty of tasks on proactive facilitation and interference. Journal of Experimental Psychology. 1953, 46, Pp. 37-42.

Bingham, W. V. Aptitudes and aptitude testing. New York: Harper & Brothers. 1937, Pp. 16-18.

Dale, H. C. A. Fault-finding in electronic equipment. Ergonomics. 1958, 1(4), Pp. 356-385.

Day, R. H. Relative task difficulty and transfer of training in skilled performance. Psychological Bulletin. 1956, 53(2), 160-167.

Department of the Air Force. Military personnel classification policy manual. AFM 35-1. August 1970.

Deupree, R. H. and Simon, J. R. Reaction time and movement time as a function of age, stimulus duration and task difficulty. Ergonomics. 1963, 6, Pp. 403-411.

Harding, F. D. and Brokaw, L. D. Implications of Air Force personnel information for job requirements. WADC-PL-TM-58-3. Lackland AFB, Texas: Personnel Laboratory, Wright Air Development Center, February 1958.

Holding, D. H. Transfer between difficult and easy tasks. British Journal of Psychology. 1962, 53(4), Pp. 397-407.

Lecznar, W. B. Three methods for estimating difficulty of job tasks. AFHRL-TR-71-30. Lackland AFB, Tex.: Personnel Research Division, Air Force Human Resources Laboratory (AFSC), July 1971.

Lindquist, E. F. Design and analysis of experiments in psychology and education. Boston: Houghton Mifflin Co. 1953, Pp 359-361.

Madden, J. M. What makes work difficult? Personnel J., Jul-Aug 1962, 41, (7), Pp. 341-344.

Mead, D. F. Continuation study on development of a method for evaluating job difficulty. AFHRL-TR-70-43. AD-720 254. Lackland AFB, Tex.: Personnel Division, Air Force Human Resources Laboratory (AFSC), November 1970.(b)

Mead, D. F. Development of an equation for evaluating job difficulty. AFHRL-TR-70-42. AD-720 253. Lackland AFB, Tex.: Personnel Division, Air Force Human Resources Laboratory (AFSC), November 1970.(a)

Mead, D. F. & Christal, R. E. Development of a constant standard weight equation for evaluating job difficulty. AFHRL-TR-70-44. AD-720 255. Lackland AFB, Tex.: Personnel Division, Air Force Human Resources Laboratory (AFSC), November 1970.

Shaw, M. E. Scaling group tasks: A method for dimensional analysis. NR 170-266, Nonr-580(11). University of Florida, Gainsville, Florida, July 1963.

Tilley, K. W. Developments in selection and training. Ergonomics. 1969, 12(4), Pp. 583-597.

Vitola, B. M. and Valentine, L. D. Assessment of Air Force accessions by draft-vulnerability category. AFHRL-TR-71-10. AD-724 094. Lackland AFB, Texas: Personnel Division, Air Force Human Resources Laboratory (AFSC), March 1971.

Warren, H. C. Dictionary of Psychology. Cambridge: Houghton Mifflin Co. 1934, 18.

PTEP-EVALUATION TECHNIQUES IN THE FBM PROGRAM

By

F. B. Braun and
B. H. Hannaford

Presented By

William Ellis

Strategic Systems Project Office
Washington, D. C.

CAT packages are distributed to training sites to be retained by the Test and Evaluation teams until required during a course of instruction. At the specified time during the course, the academic instructor receives the CAT package for same-day administration. This procedure prevents the instructor from "teaching the test." A conventional answer key is supplied to the instructor with the test package so that the test can be scored and the results used immediately by the training site. After training site use, all test results, in raw data form, are returned via the Test and Evaluation team to the central site. The test materials themselves are retained by the Test and Evaluation team for use during the next scheduled course. Normally two or more alternate forms of the test will be available for random selection.

When the raw test results are received at the central site, the material is scored and the results analyzed. A quick-look SAT report (figure 10) is generated which is then distributed to the submarine or support activity command as well as to designated higher commands. The contents of this SAT report include norms for various fleet personnel categories, individual scores separated into several knowledge and skill areas plus the overall score, and applicable training recommendations for individuals scoring below designated cut points. Applicable training includes both formal courses and self-study material. This SAT report provides personnel supervisors with a definitive tool to be used in conjunction with their personal observations for logically scheduling training.

A quick-look CAT report is not generated. Instead, a CAT summary report covering all of the tests administered during the report period is issued to the training sites as well as to designated higher commands.

Up to this point, all that has been mentioned is the use of personnel testing data. Another large portion of the PTEP consists of collecting and tracking significant personnel, equipment, and training system data to allow meaningful evaluation of the testing results. A Personnel Data Sheet (figure 11) is maintained for each individual in the program. The data sheet includes such things as a number of submarine patrols completed, previous duty assignments, watch qualifications, program entrance scores, formal courses attended, self-study materials completed, and a complete record of all CAT and SAT results. In addition, data is collected concerning the status and use of training facilities, training hardware, documentation, as well as tactical equipment status and maintenance data. This data is collected at the training sites by the Test and Evaluation Teams and from the fleet through patrol records, maintenance reports, and data information sheets. Information is also gathered from contractor reports and status summaries. The collection of this additional data, plus the personnel testing data, provides the basis for detailed analysis and evaluation of both personnel and the training system.

The size of the PTEP effort,including approximately 145 different tests covering nearly 4000 men, dictates the use of automatic data processing as an integral and continuing part of the PTEP.

ADP being implemented includes:

1. storage and update for approximately 50,000 test items
2. storage of test content parameters
3. printout of SAT and CAT knowledge section tests
4. scoring of knowledge section answer sheets and skill tests
5. processing for maintenance of test instrument characteristics
6. production of quick-look SAT reports including applicable training recommendations
7. production of CAT summary reports
8. production of Personnel Data Sheets
9. printouts of data sorts
10. query response system for special data requests

By preloading the automatic data processing system with selection and decision criteria, many of the test production and data sorting tasks can be expedited. Machine analysis printouts are used during engineering review for correlating and weighing the various data elements. Reports of deficiencies, problem areas, and recommendations are produced only after the engineering review and evaluation of the data. These reports are provided not only to higher commands to be used as system management tools, but also to applicable user commands for immediate adjustments to personnel utilization or to the training system.

Briefly then, PTEP uses the program standards, the Personnel Performance Profiles and the Training Level Assignments along with the Training Objectives Plan, in order to construct tests. The results of these tests are accumulated as part of an overall data collection task which includes collecting training system data, personnel history, and equipment data in order to provide for a meaningful evaluation for program adjustments.

The organization required to carry out the PTEP consists of four basic groups. An Evaluation Committee establishes policy for the program. This committee is composed of representatives from the Chief of Naval Operations, the Bureau of Naval Personnel, the Strategic Systems Project Office, and the submarine force commanders. An Evaluation Supervisory Group carries out the policy decisions of the Evaluation Committee, coordinates the day-to-day operation of the program, and serves as the central clearing house between the program groups and all other interested parties. The third group supporting this program is composed of the various civilian contractors who supply curricula and testing instruments, as well as the other services. As previously stated, a central site produces, packages, and distributes the tests and also coordinates the ADP and evaluation efforts for the Evaluation Supervisory Group.

Test and Evaluation Teams, composed of specifically designated personnel at each of the FBM training sites, comprise the fourth basic group. These teams brief personnel on the program, ensure proper administration of the SAT materials, coordinate CAT administration with the training site instructors, and distribute and expedite SAT reports. It can be said that in general, the Test and Evaluation Teams are the frontline troops who make the program work.

It is still too early in the evaluation phase to draw meaningful conclusions about the effectiveness of the PTEP in monitoring the Training Program. However, several submarine crews have undergone testing, quick-look SAT analysis reports including applicable training recommendations have been distributed, and other elements of the PTEP including ADP are being implemented daily. Based on the initial results and current progress, it does not seem too optimistic to predict that the PTEP goals, as well as those of the FBM Weapons System Training Program, will be achieved.

Table 127. Magnetic Disk File – Mk 88 Mods 0 and 1

ITEM NO.	KNOWLEDGE/SKILL
1.	EQUIPMENT KNOWLEDGES
1-1.	GENERAL
1-1-1.	State that the purpose of the Magnetic Disk File (MDF) is to provide mass storage of digital information of Digital Control Computer (DCC) programs.
1-1-2.	State the MDF consists of or is directly associated with the following. Include the function of each. a. Disk drive unit – electromechanical interface between the MDF logic and the disks b. Disk drive electronics – converts signals to and from the disk drive unit and controller into levels usable by both c. MDF controller – logic link between the disk drive unit and the DCC for both control signals and digital data transfers d. Power protection panel – controls the application of power to the MDF e. MDF patch panel – allows connecting of either MDF to either DCC
1-1-3.	Define the abbreviations, terms, and symbols used with the MDF (for example, CAR, EOC, Sync Bit).
1-1-4.	State the operational characteristics and capabilities of the MDF. a. Interprets and executes 16 instructions b. Rotational speed of 1500 rpm c. Disk pack capacity of 55-million bits d. Uses double frequency and nonreturn-to-zero recording techniques e. Has four basic operating modes for performing data transfer instructions (1) Sector (2) Track (3) Sector cylinder (4) Track cylinder
1-2.	PHYSICAL DESCRIPTION
1-2-1.	Describe all major and associated components of the MDF. Include name, nomenclature, physical appearance, reference designator, location, and construction features. a. Disk drive unit (1) Main casting (2) Hydraulic actuator (3) Carriage housing assembly (4) Spindle (5) Hydraulic pump (6) Data disk pack

FIGURE 1

Table 127. Magnetic Disk File – Mk 88 Mods 0 and 1 (Continued)

ITEM NO.	KNOWLEDGE/SKILL
2.	EQUIPMENT SKILLS
2-1.	OPERATION
2-1-1.	Perform operating procedures for the MDF.
2-1-2.	Adhere to personnel and equipment safety precautions during operation.
2-2.	MAINTENANCE
2-2-1.	Use special tools and test equipment required for maintenance of the MDF, as prescribed in applicable documentation.
2-2-2.	Perform preventive maintenance procedures on the MDF as scheduled by PMMP and presented in SMP.
2-2-3.	Perform alignment, adjustment, and calibration procedures.
2-2-4.	Perform operational tests for maintenance.
2-2-5.	Recognize and interpret indications of malfunctions.
2-2-6.	Perform documented procedures for systematic fault isolation.
2-2-7.	Use improvised procedures to isolate faults which cannot be located with documented procedures.
2-2-8.	Disassemble, repair, and reassemble the MDF to the authorized maintenance level.
2-2-9.	Adhere to personnel and equipment safety precautions when performing maintenance.
2-3.	INSTALLATION
2-3-1.	Initial installation of equipment comprising fire control system will be accomplished by shipyard personnel. Once installed, the system itself will not be removed or replaced until the ship returns to the shipyard for overhaul. Therefore, tender and SSBN personnel will only be concerned with removal and replacement of equipment components. However, the Disk Drive Unit (DDU) and Data Disk Pack are removable and replaceable as an entire unit.
2-3-2.	Unpack and visually inspect a DDU and a data disk pack for shipping damage and ensure that all applicable hardware and/or software are available.

FIGURE 2

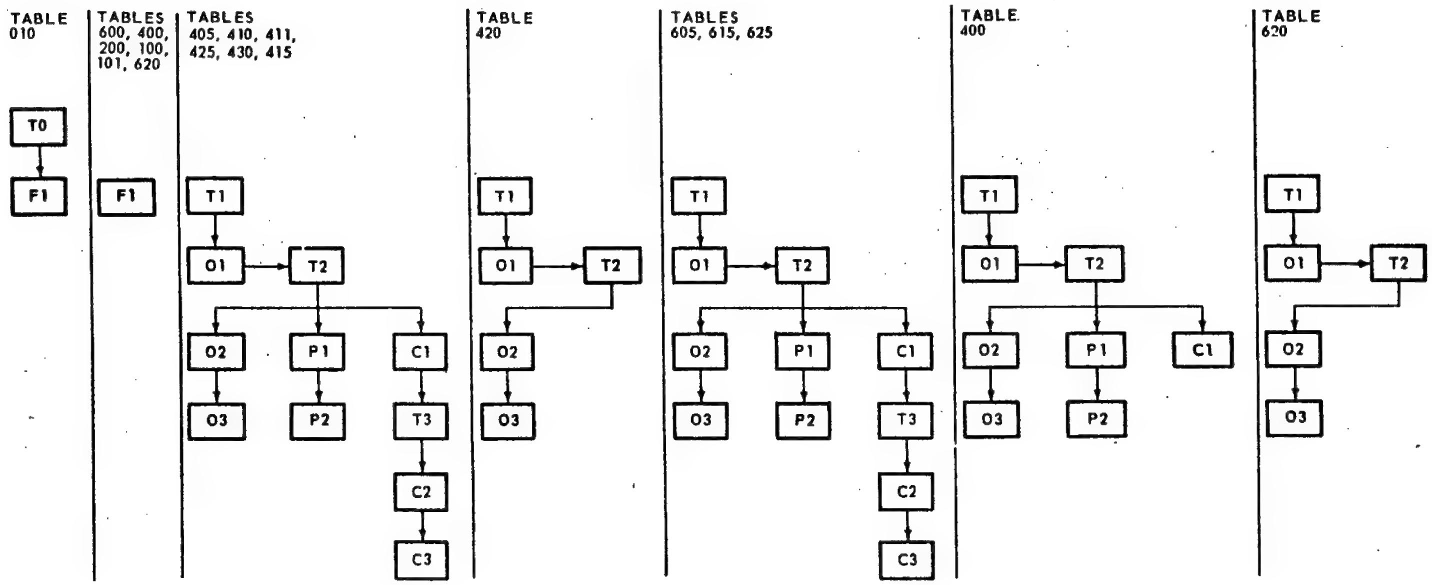


TABLE INDEX

010	FBM Weapon System (System Level)
100	Fire Control Systems - Mk 88 Mods 0 and 1 and Mk 84 Mod 1 (Subsystem Level)
101	Fire Control System - Mk 80 Mod 2 (Subsystem Level)
200	Missile Systems (Subsystem Level)
400	Missile Launching Systems (Subsystem Level)
405	Missile Launcher Tube Groups Mk 12 Mod 0, Mk 7 Mods 0 and 1, Mk 9 Mod 0, and Related Equipment
410	Missile Ejector Groups Mk 7 Mod 0 and Mk 8 Mod 0
411	Missile Ejector Group Mk 4 Mod 0
415	Missile Launcher Control Groups Mk 18 Mods 0, 1, and 2, and Mk 28 Mod 0
420	Missile Handling Equipment Mk 4 Mods 0 and 1 and Mk 5 Mod 0
425	Missile Launcher Pressurization System
430	Missile Launcher Hydraulic System
600	Weapon System Ship Support Subsystem (Subsystem Level)
605	Dehumidification System
615	Missile Heating, Cooling, and Temperature Control System
620	Electrical System
625	Missile Compensation System

FIGURE 3

Training Level Assignments for Fire Control Technician (SSBN)
Poseidon, FCS Mk 88 (NEC FT-3306)

TABLE 127

ITEM	LEVEL											
	F1	T1	T2	T3	O1	O2	O3	P1	P2	C1	C2	C3
1-1-1		x										
1-1-2		x	x									
1-1-3		x	x	x								
1-1-4		x	x									
1-2-1		x	x	x								
1-2-2		x										
1-3-1		x	x	x								
1-3-2		x	x	x								
1-3-3		x	x	x								
1-4-1		x	x	x								
1-4-2		x	x	x								
1-5-1		x	x									
1-5-2		x	x									
1-5-3		x	x									
1-5-4		x										
1-6-1			x									
1-6-2			x	x								
1-6-3			x									
1-6-4			x	x								
1-6-5			x	x								
1-6-6			x	x								
1-6-7			x	x								
1-6-8			x	x								
1-6-9			x	x								
1-6-10			x									
1-6-11		x	x	x								
2-1-1					x	x	x					
2-1-2					x							
2-2-1								x	x	x	x	x
2-2-2								x	x			
2-2-3								x	x	x	x	x
2-2-4								x	x	x	x	x
2-2-5										x	x	x
2-2-6										x	x	x
2-2-7										x	x	x
2-2-8										x	x	x
2-2-9								x		x		
2-3-1											x	x
2-3-2											x	x
2-3-3											x	x
2-3-4											x	x
2-3-5											x	x
2-3-6											x	x
2-3-7											x	x
2-3-8											x	
	F1	T1	T2	T3	O1	O2	O3	P1	P2	C1	C2	C3

FIGURE 4

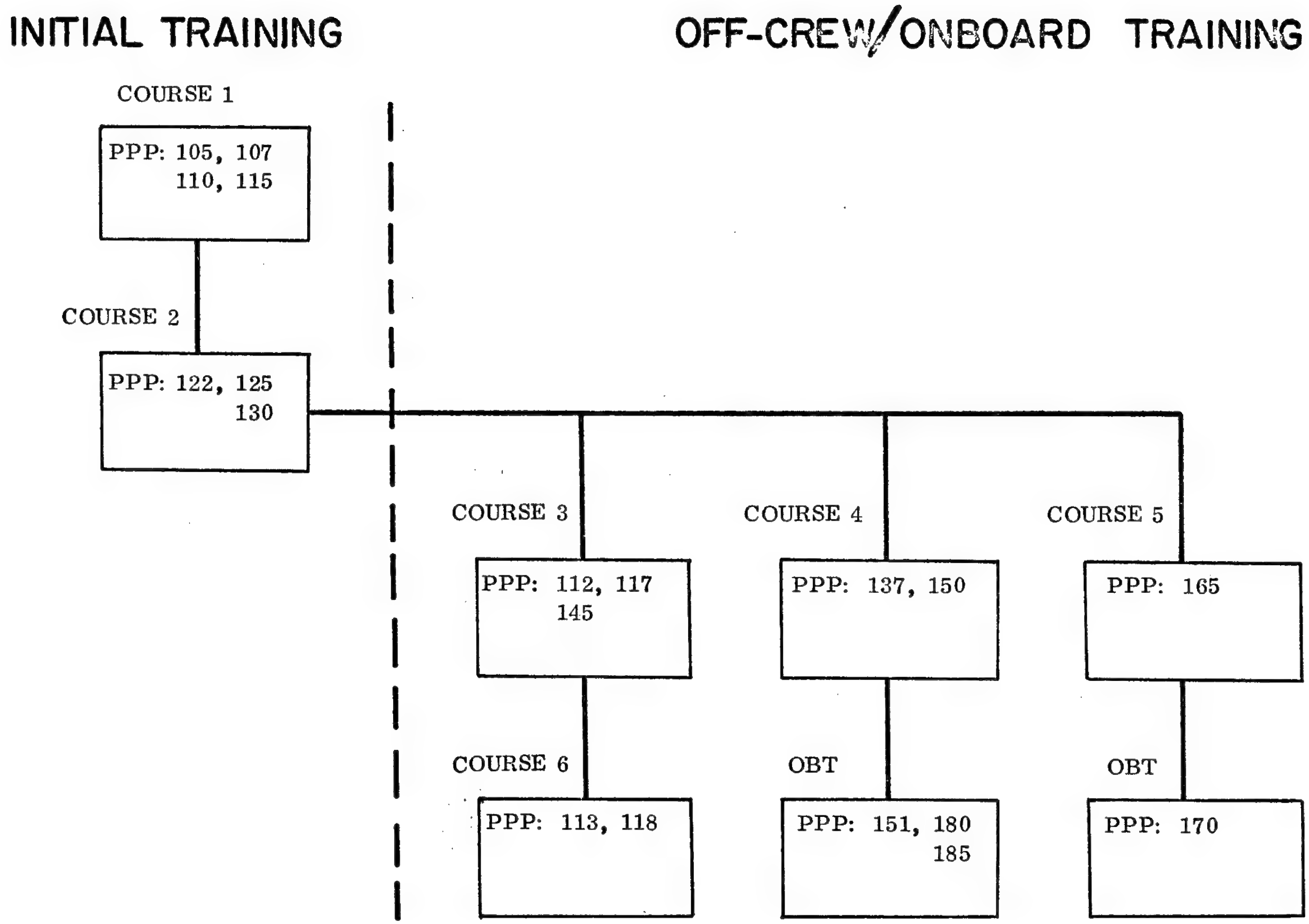


FIGURE 5 PPP TABLE ASSIGNMENT CHART

MCC SUPERVISOR WATCH QUALIFICATION GUIDE (MK 88)

NAME______________________ RATE_____ DATE ASSIGNED_________

DATE TO QUALIFY________

DATE QUALIFIED_________

I. PREREQUISITES:

1. Completed MCC Technician Qualification ________ Date

II KNOWLEDGE REQUIREMENTS:

1. Explain the purpose, organization, content and utilization of the following:

A. OP 3586/OP3804 ________

B. NAVORD OD 43144 VOL 1, VOL 2, and VOL 3 (Applicable PG's and WP's) ________

C. Applicable SMP's ________

D. Stoppage Reports ________

E. SPALT Records ________

F. MCC/Fire Control/MTRE Work Log ________

G. FCD'S ________

2. Explain the purpose, function and operation of the following, using appropriate documentation, to schematic diagram level:

A. MK 88 FCS:

(1) Test Subsystem (TESS) ________

FIGURE 6

IV EXAMINATIONS:

1. Satisfactorily completed MCC walkthrough

_____________ Date

MCC Supervisor _____________

2. Satisfactorily completed written examination (optional)

_____________ Date

3. Examined and recommended for qualification by a board consisting of:

_____________ Date

MCC SUPERVISOR	MCC SUPERVISOR	WEPS/A WEPS

4. Interviewed and recommended for qualification as an MCC Supervisor Watchstander.

_____________ Date

_____________________________ Weapons Officer

5. Interviewed and Qualified as an MCC Supervisor Watchstander.

_____________ Date

_____________________________ Commanding Officer

FIGURE 7

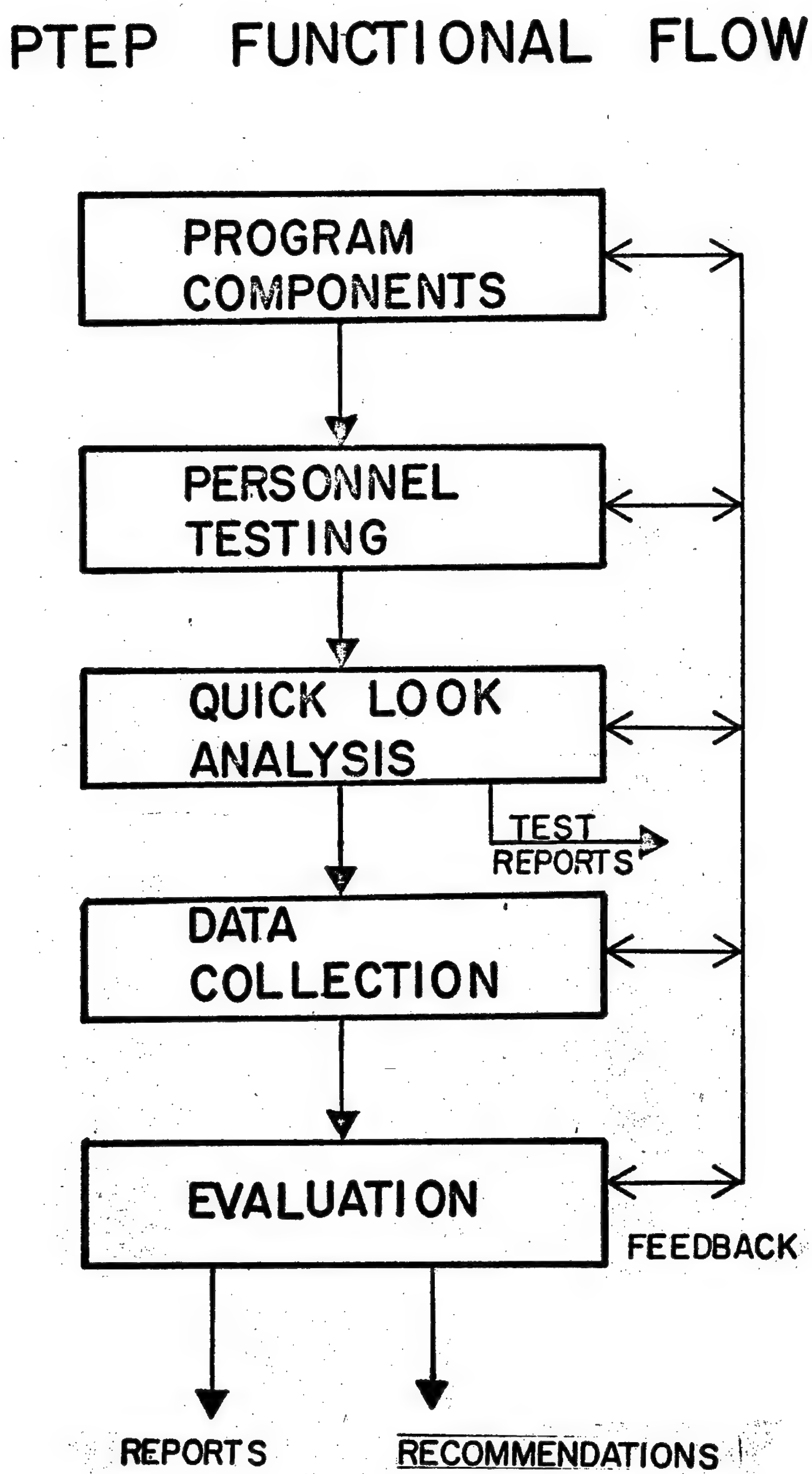

PTEP FUNCTIONAL FLOW
PROGRAM COMPONENTS
PERSONNEL TESTING
QUICK LOOK ANALYSIS
TEST REPORTS
DATA COLLECTION
EVALUATION
FEEDBACK
REPORTS
RECOMMENDATIONS

FIGURE 8

CONSTRUCTION OF PTEP TESTS

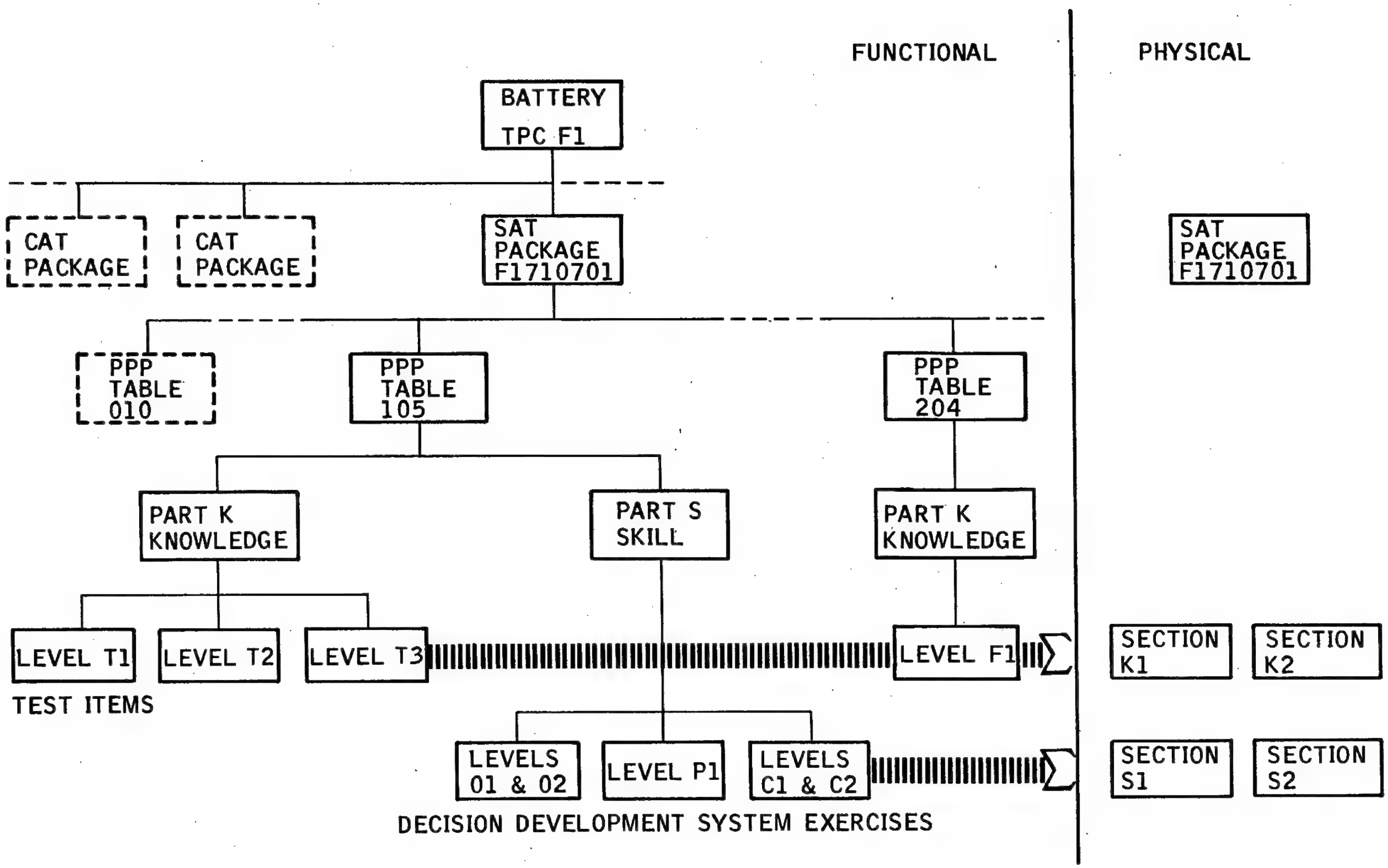


FIGURE 9

INDIVIDUAL SAT REPORT

SSBN 695 BLUE

NEC 3325

TEST NUMBER NS1710701

TEST DATE 71-06-22

AREA DESCRIPTION CODES

KNOWLEDGE AREAS

A SINS MK 2 MOD 6
B MARDAN AND TEST SET
C MULTISPEED REPEATER
D COMPUTER AND NAVAIDS
E OPTICAL EQUIPMENT
F SELECTRIC TYPEWRITER
G CHECKOUT AND TEST EQUIPMENT
H SUBSYSTEMS

SKILL AREAS

I SINS DDS
J MARDAN DDS
K MULTISPEED REPEATER DDS

NEC KNOWLEDGE AVERAGES

1.	FLEET	165
	A. E7/E8	175
	B. E6	165
	C. E5	158
	D. E3/E4	124
2.	CREW	168

	KNOW. SCORE	A	B	C	D	E	F	G	H	I	J	K
		KNOWLEDGE AREAS								SKILL AREAS		
FLEET NEC AVERAGES	165	13	24	27	7	36	32	11	13	240	320	160
HENRY, PRINCE ET1 758-21-3694	155	17	18	35	10	22	25	13	15	245	300	170

APPLICABLE TRAINING AREA		DDS	SSWB	COURSES
B	MARDAN AND TEST SET	N10	N16	F 193-018
E	OPTICAL EQUIPMENT	N8	N13, N17	F 193-085, F 193-020
F	SELECTRIC TYPEWRITER	N12	N5, N22	F 193-077

FIGURE 10

PTEP PERSONNEL DATA

NAME 001 | SSAN 002 | RATE AND DOR 003 | PRATE AND DOR 004 | PNEC AND DATE 005 | SNEC AND DATE 005

TNEC AND DATE 007 | TPC 009 | POL PAT 010 | POS PAT 011 | DUTY STA AND DROB 012 | INST EXP MO 013 | CONST/CONV MO 014 | CIV ED YR 015

PRD 016	EAOS 017	SSQUAL/REQUAL 018	DUTY STA HISTORY 030/034	BASIC BATTERY SCORES 035/047		CORR/OFF DUTY CS AND DOC 050/054
			030	035	AFQT	050
			031	036	GCT	051
			032	037	ARI	052
			033	038	MECH	053
			034	039	CLER	
				040	SPT	
				041	SONAR	
				042	RADIO	
				043	ETST	
				044	READ	
				045	MATH	
				046	PHYS	
				047	ELEC	

COURSE COMP AND DOC 055/064	SSWB COMP 065/084					CAT RECORD 100/119				SAT RECORD 120/129			
		SSWB	VOL	PART	DOC		TEST NO	DATE	SCORE		TEST NO	DATE	SCORE
055	065					100				120			
056	066					101				121			
057	067					102				122			
058	068					103				123			
059	069					104				124			
060	070					105				125			
061	071					106				126			
062	072					107				127			
063	073					108				128			
064	074					109				129			
	075					110							
	076					111							
	077					112							
	078					113							
	079					114							
	080					115							
	081					116							
	082					117							
	083					118							
	084					119							

FIGURE 11

by F. B. Braun and B. H. Hannaford, Data-Design Laboratories

The Fleet Ballistic Missile (FBM) Weapons System, which encompasses both the Polaris and the Poseidon missiles and their support, has a record of maintaining a high degree of readiness. This record has been achieved not only by procurement of highly reliable hardware but also by initially training and maintaining a personnel force of well qualified, highly skilled technicians and officers.

Experience gained over the years with Polaris training in producing these highly skilled personnel led to certain observations which made it appear desirable to restructure various aspects of the program for the Poseidon training effort. These observations were:

1. The training pipeline was too long- in excess of twelve months for most technicians- too expensive, and too often led to a highly trained civilian TV serviceman or electronics technician after completion of initial obligated service.

2. No means existed to adequately measure technician capability outside the formal school environment.

3. No way, other than personal observation, existed to determine which man should be called upon to fix which equipment.

4. Training conducted between patrols was often less than satisfactory because of a lack of information as to exactly which courses would be of the most benefit.

In order to provide a systematic management approach towards solving these problems the Chief of Naval Operations established, through a formal instruction, the FBM Weapons System Training Program. This training program has provisions for:

1. Accurate, current, and comprehensive job-related training.

2. Identification and correction of training problems.

3. Effective utilization of training resources.

4. Evaluation and feedback of training results.

The underlying purpose of the Training Program is the restructuring of the training system to use minimum training resources and still produce personnel sufficiently trained and qualified to support the high degree of fleet operational readiness.

The four basic components of the FBM Weapons System Training Program are:

1. Personnel Performance Profiles (PPP)

2. Training Path Charts (TPC) and Training Objectives Plan (TOP)

3. Personnel Qualification Guides (PQG)

4. Personnel and Training Evaluation Plan (PTEP)

The subject of this paper is the Personnel and Training Evaluation Plan: the evaluation portion of the Training Program and the means by which adjustments may be made to the program. In order to put the PTEP in the proper perspective, I would like first to discuss the other components of the overall Training Program.

Comprehensive listings of the skills and knowledges required to operate and maintain the various equipment of the FBM Weapons System were first developed to provide a standard, or baseline, which could be used to measure capability to support the equipment. These listings, entitled Personnel Performance Profiles (PPP), were prepared for individual equipment, subsystems, or systems, as well as for background material.

The Profile Tables list requisite knowledges and skills in these categories:

Knowledges

1-1 General
1-2 Physical Description
1-3 Functional Description
1-4 Interface Description
1-5 Operational Description
1-6 Maintenance Description and Documentation

Skills

2-1 Operation
2-2 Maintenance
2-3 Installation

A typical PPP table is shown in part on figures 1 and 2.

After the Personnel Performance Profiles were prepared listing the skills and knowledges required by the hardware, personnel responsibilities for these skills and knowledges were assigned. This was accomplished by part of the next component of the Training Program: The Training Path Chart (TPC) (Figure 3). The skill and knowledge levels are depicted on the chart by blocks. The T blocks are Theory, the O blocks are Operation, the P blocks are Preventive Maintenance, and the C blocks are Corrective Maintenance. The numeral portions of the alphanumeric block designators represent achievement levels. The entire chart represents the desired level of

attainment, by PPP table, for the specified fleet personnel. The chart thus covers not only initial training but also the advanced skills and knowledges taught at the FBM training sites, learned during onboard training, or gained through experience. A Training Level Assignment chart (figure 4) accompanies the Training Path Chart and identifies by achievement level the skill and knowledge items from each PPP table for which the specified individual is responsible.

The Personnel Performance Profiles, Training Path Charts and Training Level Assignments thus provide the necessary standards for the FBM Weapons System. These standards are updated and maintained current as equipment is replaced or altered or as the operational requirements of the fleet change.

The Training Objectives Plan (TOP) completes the second basic component of the FBM Weapons System Training Program. The TOP has as its goal the establishment of both formal and informal courses to support the desired achievement levels designated by the standards. Formal off-crew training courses, which are conducted during the between-patrol period, as well as formal initial training courses require the acquisition of curricula. These curricula are constructed using a published guideline which specifies how the standards are to be used in establishing course content, depth of coverage, and sequence of presentation. The PPP Table Assignment Chart (figure 5) illustrates how curricula are keyed to the system standards to facilitate update, testing, and evaluation.

The TOP has provisions for establishing informal courses for those training areas where it is impractical or inefficient to establish formal courses. These informal courses consist of self-study material and are used by fleet personnel during both on-board and off-crew periods. These informal courses are of particular benefit to submarine tender and other support personnel who normally can not attend formal courses after their initial training.

The next component of the Training Program is the establishment of a standardized set of watch qualification and maintenance requirements. These requirements, entitled the Personnel Qualification Guides (PQG), list watch station and maintenance qualifications necessary to operate and maintain the FBM Weapons System. The PQG provide a uniform system whereby a command can determine approximate qualification levels and capabilities of personnel transferred from other commands within the program. Figures 6 and 7 are an example of part of a PQG.

Thus far, we have seen that the FBM Weapons System Training Program consist of:

1. the standards- Personnel Performance Profiles, Training Path Charts, and Training Level Assignments

2. the methodology of training- the Training Objectives Plan

3. the means of maintaining uniformity of qualification in the fleet- the Personnel Qualification Guides.

The final component of the FBM Weapon System Training Program, the Personnel and Training Evaluation Plan (PTEP) is the means by which the effectiveness of the entire program is measured. The PTEP has as its responsibilities the qualitative assessment of personnel technical proficiency and the adequacy of training. The PTEP will provide reports to system managers to assist them in making intelligent, knowledgeable decisions. These reports will cover areas such as personnel, training facilities, hardware, documentation, and courses of instruction.

The PTEP, as shown in the functional flow (figure 8), consists of several elements already implemented or being implemented over the next few months.

Personnel testing consists of a battery, or series, of tests administered to personnel from the time they enter the Training Program until the time they leave the system. The tests are all based on the program standards and measure both course of instruction comprehension and total skill and knowledge achievement.

Quick look analysis results of personnel testing are provided to individual commands to assist in effective scheduling of both formal and informal personnel training.

Further analysis and evaluation of personnel testing results are performed after the input of additional pertinent data such as personnel history, equipment status, documentation status, and training facility information. This data collection effort and the subsequent evaluation of the data provide the basis for reports and recommendations which then close the loop for the FBM Weapons System Training Program.

Up to this point, the system which is to be evaluated and the goals and purposes of the evaluation have been described. I would now like to discuss in greater detail the techniques, the methods, and the testing materials used in the PTEP.

By far the largest task in terms of required materials is personnel testing. Two types of tests are used:

1. A System Achievement Test (SAT) is designed to measure the capability of fleet personnel on the overall skills and knowledges designated by the standards. This comprehensive sampling of total technical skill and knowledge requirements is administered at regular intervals to submarine and support personnel.

2. A Course Achievement Test (CAT) is designed to measure the level of comprehension of personnel taking formal training courses. A CAT consists of test instruments covering the Profile items for which the curriculum was constructed. Thus, a CAT checks the adequacy of the curriculum and related training support material as well as personnel comprehension.

System Achievement Tests and Course Achievement Tests are both constructed to measure skills and knowledges. The Personnel Performance Profiles and the Training Level Assignments are used for both tests as the basis for determining test content. Construction of PTEP tests (figure 9) is accomplished along the following general lines.

A battery of tests is designated which includes both CAT and SAT packages as individually administered tests. The battery covers the skills and knowledges for specific personnel from the time they enter the program, through initial and off-crew training, and throughout fleet operation. Each package, whether a SAT or a CAT, is developed to cover all of the Profile material designated by the applicable Training Level Assignment. The Profile material is covered by developing two parts for the package - a knowledge part and a skill part. Both parts contain test instruments to test sufficiently an individual's knowledge or skill capability to the applicable achievement levels.

For the knowledge parts, the test instruments used are multiple choice test items, either open book or closed book depending upon subject matter and operational requirements. These test items are procured from hardware contractors and from Navy training sites. All of the test items are written in accordance with a detailed specification to ensure conformity to the standards of the Training Program as well as uniformity of format.

For the skill parts, the test instruments currently being used are Decision Development System exercises. These exercises are paper and pencil simulations of equipment and utilize erasable ink techniques. A paper presented at last year's meeting of the Military Testing Association discussed the feasibility of using Decision Development System material as a testing device. These skill test materials, like the knowledge test items, are obtained from a civilian contractor and are keyed to the program standards.

The test instruments are selected for a particular test, SAT or CAT, utilizing a sampling procedure and a test content plan developed from the Personnel Performance Profiles and Training Level Assignments. After the test instruments have been selected, they are arranged by similar types into physical test sections for ease of administration. Each test section is a physical entity, can be administered as such, and requires certain support materials. For example, a typical knowledge section consists of a Test Booklet containing the multiple choice test items, an Illustration Booklet if required by the test items, machine-scoreable answer sheets, and a Proctor Record Sheet for recording pertinent data regarding the group of examinees. In addition, a Proctor Guide is provided for each section to ensure consistency and accuracy in administration.

All of the required materials for a test package are assembled at a central site. The materials are packaged by physical section for transmittal to the testing site. Each package contains complete instructions for the use of the enclosed materials, directions concerning materials which must be supplied locally at the site, and instructions for the return of the materials to the central site.

SAT packages for submarine personnel are distributed to off-crew training sites and upon receipt are administered by special Navy Test and Evaluation teams. SAT packages for support personnel are delivered to the support activity and upon receipt are administered by local personnel designated for this function. Upon completion of the test, all materials, in raw data form, are returned to the central site for processing.

A VALIDITY ASSESSMENT OF THE NAVAL ADVANCEMENT EXAMINATIONS THROUGH MULTIPLE DISCRIMINANT FUNCTIONS

CASIMER S. WINIEWICZ
U. S. NAVAL EXAMINING CENTER

During a typical year approximately 1,700 different examinations are developed at the U. S. Naval Examining Center to reflect the scores of technical skills and fields of knowledge required in today's Navy. These tests are administered semi-annually in February and August under completely controlled conditions. This entails shipping and accounting for three-quarters of a million examinations, as well as processing their results and promulgating the information back to the examinees. It takes less than 60 days from the time a candidate takes an examination until he receives the results.

In order for a candidate to advance in rate (occupation), he must be qualified in all respects, recommended by his Commanding Officer and successfully compete on a service wide advancement examination. Each examination is weighted to determine the number of items which are to be included for each subject matter area breakdown under the applicable job level for the rate. Each instrument at Pay Grades E-6 and E-7 is composed of two major parts: professional which pertains to the specific job area (usually 125 - 4 option multiple-choice items) and the military (usually 25 - 4 option multiple-choice items). Pay Grades E-4 and E-5 have a completely professional examination. Because of the lack of a rigidly standardized examination (as contrasted with commercial instruments) and the fact that each examination is utilized only once on a given population to preclude compromise or collusion, and also to maintain the flexibility of the

instrument to reflect current technological changes, some control factor is required to tie the test into previous test results to insure comparable standards. Pre-testing is not feasible because of the magnitude of the overall program and the relatively tight scheduling required in the continual development and processing of all examinations.

The control factor that is utilized to tie in present examinations with previous results is to control the design of each new examination by including a minimum of 50% of the items previously used with known statistical characteristics obtained through item analysis, and a constant check of separate periodic research studies on the examinee population. The remaining 50% of the items are newly developed for use in the current examination. Consequently through a rigid control of item and statistical analyses and related special studies, the test instrument assumes greater standardization and control than some of our so called standardized commercial instruments on the market today.

In determining an equitable cut-off score for each of the 1700 different examination populations that will dichotomize our groups into pass - fail categories and still maintain comparability between different examinations and examining periods, a multitude of factors must be taken into consideration, such as needs of the service at large, budgetary problems, and above all the qualifications and performance of each candidate. The raw scores are converted to standard scores (via linear and non-linear transmutations) with a range from 20 - 80, with a mean

of 50 and a sigma of 10. This scale coincides with t scores and takes advantage of all the properties of the normal curve.

Final advancement, assuming the candidate has fulfilled all the necessary prerequisite requirements and has attained a passing score on the examination is determined by one of two methods depending upon the criticality of the occupation under consideration. First if the candidate happens to compete in an occupation whose skills are in demand (more vacancies available than qualified personnel to fill them), the examination passing score qualifies the candidate for advancement. However, if the candidate competes in an occupation where the number of vacancies are small (more qualified personnel available than existing vacancies), final selection of those authorized to be advanced is made on the basis of relative standing on a final composite score consisting of the following five factors, along with their maximum values: (1) Performance Factor (50), (2) Length of Service (20), (3) Time in Rate (20), (4) Number of Awards (15), and (5) Examination Grade (80), giving a maximum composite score of 185.

This system provides an equitable opportunity to compete for advancement, under completely controlled test conditions for the number of authorized advancements regardless of the place that an individual may be stationed throughout the world, his present duty assignment, or the vacancies or surpluses that exist in the local commands. The Navy maintains approximately 3200 activities (ships or stations) located in all parts of the world, and classifies men into over 65 different occupational skills divided into 6 different levels of ability.

It is the function of the U. S. Naval Examining Center within this framework to; construct examinations, ship and receive them back for processing, account for every examination used or unused, evaluate statistically all examinations, maintain the integrity of the examining system, select the most qualified candidates for advancement, and finally to continually strive to improve the entire system.

The Chief of Naval Personnel assigned to the Naval Examining Center, the responsibility of ascertaining the validity of examinations utilized in the Navy Wide Advancement System.

The August 1970 (Series 55) Navy Wide Advancement Examination population was decided upon as the population to be included for the validity study. During this examining cycle period, roughly 150,000 candidates competed in Pay Grades E-4 through E-7 for advancement in 369 different rates. The number of usable returns from this population should amount to approximately 100,000 candidates.

A total of twenty-eight primary variables are utilized in the overall analysis. Twenty variables are descriptive and eight are utilized as classifier variables.

The following 14 tables are descriptive of the approach utilized in the study. Tables 1 through 7 are descriptive in nature, and the remaining tables are based on data computed for a single rate.

Table 1 - Refers to the evaluation form that was utilized to collect data from the field, in a manner that would elicit cooperation and not interfere too much in the respective activities' day to day operation.

The form was designed as an answer sheet that required the respondee to only blacken certain pertinent information and could be optically scanned when returned to NEC. The sheet was divided into 3 main sections. Section 1 is for NEC use only and contains the pre-printed information on the candidate to be evaluated. Section 2 requires the candidates immediate supervisor to only blacken 3 circles of his choice. The remaining section is filled in by the administrative office of the activity and only requires 8 appropriate circles to be blackened out.

Table 2 - Defines the August 1970 (Series 55) Navy Wide Advancement Examination population that was utilized for the validity study. During this examining cycle period, roughly 150,000 candidates competed in Pay Grades E-4 through E-7 for advancement in 369 different rates.

The number of usable returns from this population should amount to approximately 100,000 candidates. This figure allows for 2/3 of the original population to be included for analysis in the validity study.

Table 3 - Illustrates the twenty-eight variables that have been selected for inclusion in this study.

Table 4 - Defines the eight classifier variables and the manner in which they are sub-divided.

Table 5 - Illustrates the type of proportionate weighting data that will be generated for all occupations.

Table 6 - Is an example of the scattegrams that will be produced for all occupations and variables.

Table 7 - Illustrates the general approach that will be utilized in producing the data for the study. Because of the combination and permutations of variables, literally hundreds of thousands of documents can be produced.

Table 8 - Correlation matrix for the 28 variables calculated for the YN2 rate.

Table 9 - Correlation matrix for YN2 Class A school graduates.

Table 10 - Correlation matrix for YN2 non-school population.

Table 11 - Scattegram with regression equations for YN2 Class A school population.

Table 12 - Scattegram with regression equations for YN2 non-school population.

Table 13 - Correlation matrix for YN2 candidates working within their occupation.

Table 14 - Correlation matrix for YN2 candidates working outside of their occupation.

TABLE 1

(Name) (Serial No.) (Rate) (Activity Code) (Cycle) (Exam. Serial)

SECTION 1 SHOULD BE COMPLETED FIRST BY THE RATER

Section 1: Instructions to the Rater

1. The rater should be above named individual's immediate supervisor. Evaluations will be used for research purposes only.
2. Evaluation should not be influenced by the individual's last performance evaluation or his performance on the last advancement examination.
3. Based on your observation of his daily performance of duties rank the individual in comparison with all other naval personnel of his rating and paygrade.
4. Consider only such factors as: Knowledge and skills acquired, and ability to perform assigned duties. Do not consider factors such as: loyalty, leadership, conduct or personality traits which are not measured by the advancement examinations.

RATING SCALE

	UPPER					LOWER				
Interpretation:	10%	20%	30%	40%	50%	50%	40%	30%	20%	10%
Value:	1	2	3	4	5	6	7	8	9	10

5. No rater should evaluate more than 10 people: Total number of people rated:
① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨ ⑩
6. Each candidate besides being rated should be assigned a relative rank order number among his group. (i.e.) 5 people rated — numbers for each candidate should read 1 through 5. (No ties) Below check rank order number assigned for this candidate:
① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨ ⑩

SECTIONS 2 AND 3 SHOULD BE COMPLETED BY THE ADMINISTRATIVE OFFICE AFTER SECTION 1 IS COMPLETED

Section 2: Candidates last performance rating

Range: From / To	4.00 / 3.81	3.80 / 3.61	3.60 / 3.41	3.40 / 3.21	3.20 / 3.01	3.00 / 2.81	2.80 / 2.61	2.60 / 2.41	2.40 / 2.21	2.20 / Below
1. Professional Perf:	O	O	O	O	O	O	O	O	O	O
2. Military Behavior:	O	O	O	O	O	O	O	O	O	O
3. Leadership:	O	O	O	O	O	O	O	O	O	O
4. Military Appearance:	O	O	O	O	O	O	O	O	O	O
5. Adaptability:	O	O	O	O	O	O	O	O	O	O

Section 3: Additional Information

1. From what type(s) of service school did candidate graduate pertaining to his rating:
Class: A B. C None
2. Candidate is currently working In or Out of his rating:
3. Candidate is Career, Non-Career or Undecided:

NEC VALIDITY STUDY

Return All Completed Forms To:

COMMANDING OFFICER (Code 00)
U. S. NAVAL EXAMINING CENTER
BUILDING 2711
GREAT LAKES, ILLINOIS 60088

EXAMPLE

WRONG

RIGHT

TABLE 2

SERIES 55 FIGURES

PAY GRADE	NO. RATES	SHIP-N*	EXP. N
E-4	99	51,762	36,000
E-5	96	36,289	25,000
E-6	86	23,656	16,000
E-7	88	34,501	25,000
TOTALS	369	146,408	102,000

* CHECKED BY ACTIVITY UTILIZATION LISTING

TABLE 3

NEC VALIDITY STUDY
PRIMARY VARIABLES

	1. RATING SCALE VALUE
	2. RELATIVE RANK
	3. WEIGHTED POSITION-RANK-EVALUATION FACTOR
ENL. PERF. MARK	4. PROFESSIONAL PERFORMANCE
	5. MILITARY BEHAVIOR
	6. LEADERSHIP & SUPERVISORY ABILITY
	7. MILITARY APPEARANCE
	8. ADAPTABILITY
	9. TOTAL PERFORMANCE EVALUATION MARK (4+5+6+7+8)
FINAL MULT.	10. EXAMINATION STANDARD SCORE
	11. LENGTH OF SERVICE (LOS)
	12. TIME IN RATE (TIR)
	13. AWARDS
	14. FINAL MULTIPLE (9+10+11+12+13)
	15. EXAMINATION RAW SCORE
BASIC BATT.	16. GCT
	17. ARI
	18. CLER
	19. MECH
	20. GCT + ARI
CLASSIFIER VAR.	21. SERVICE SCHOOL(S)
	22. WORKING IN/OUT OF RATE
	23. CAREER STATUS
	24. USN/R
	25. SEX
	26. RACE
	27. EDUCATIONAL LEVEL
	28. PROCESSING CODE

TABLE 4

CLASSIFIER VARIABLE BREAKDOWN

SERVICE SCHOOLS

1. A SCHOOL
2. B SCHOOL
3. C SCHOOL
4. NONE

WORKING/RATE

1. IN RATE
2. OUT OF RATE

CAREER STATUS

1. CAREER
2. NON-CAREER
3. UNDECIDED

USN/R

1. USN
2. USNR

SEX

1. MALE
2. FEMALE

RACE

1. CAUCASION
2. NEGROID
3. INDIAN(A) & MONGOLIAN
4. MALAYAN

EDUCATIONAL LEVEL

1. 8 YRS OR LESS
2. 9 YRS
3. 10 YRS
4. 11 YRS
5. 12 YRS
6. 12 YRS-H.S. DIP.
7. X YRS-GED
8. 13 YRS
9. 13 YRS-1 YR EQIV.
10. 14 YRS
11. 14 YRS-ASSOCIATE D.
12. 16 YRS-B.A.
13. 17 YRS-M.A.
14. 18 YRS-M.A.

PROCESSING CODE

1. ADVANCE
2. PNA
3. FAIL

TABLE 5

PROPORTIONATE WEIGHT BREAKDOWNS

FINAL MULTIPLE WEIGHTS

	PER FACT	LOS	TIR	AWD	SS
Remainder	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX
Part Remainder	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX
Beta Wts.	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX
Sect. Wts.	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX
Means	XX.XX	XX.XX	XX.XX	XX.XX	XX.XX
Std. Dev.	XX.XX	XX.XX	XX.XX	XX.XX	XX.XX

PERFORMANCE FACTOR SCALE WEIGHTS

	PRO. PER.	MIL. BEH	LED.	MIL.ADAP.	ADAP.
Remainder	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX
Part Remainder	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX
Beta Wts.	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX
Sect. Wts.	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX
Means	XX.XX	XX.XX	XX.XX	XX.XX	XX.XX
Std. Dev.	XX.XX	XX.XX	XX.XX	XX.XX	XX.XX

TABLE 6

RATE:

FACTOR: GCT + ARI

POPULATION: CLASS A SCHOOL GRADS

N = 182

EXAM CUT SCORE

GCT + ARI CUT

SS SCORE \ FACTOR	[illegible]	61 / 9 / 70	71 / 8 / 80	81 / 7 / 90	91 / 6 / 100	101 / 5 / 110	111 / 4 / 120	121 / 3 / 130	131 / 2 / 140	141 / 1 / 160
76 – 80										
71 – 75									3	
66 – 70							1	3	6	2
61 – 65							1	6	10	5
56 – 60							2	9	7	
51 – 55					1	1	7	16	12	1
46 – 50					2	3	11	21	6	
41 – 45						2	8	8	3	
36 – 40						1	6	3	4	
31 – 35						1	3	4		
26 – 30						2	1			
21 – 25										

TABLE 7

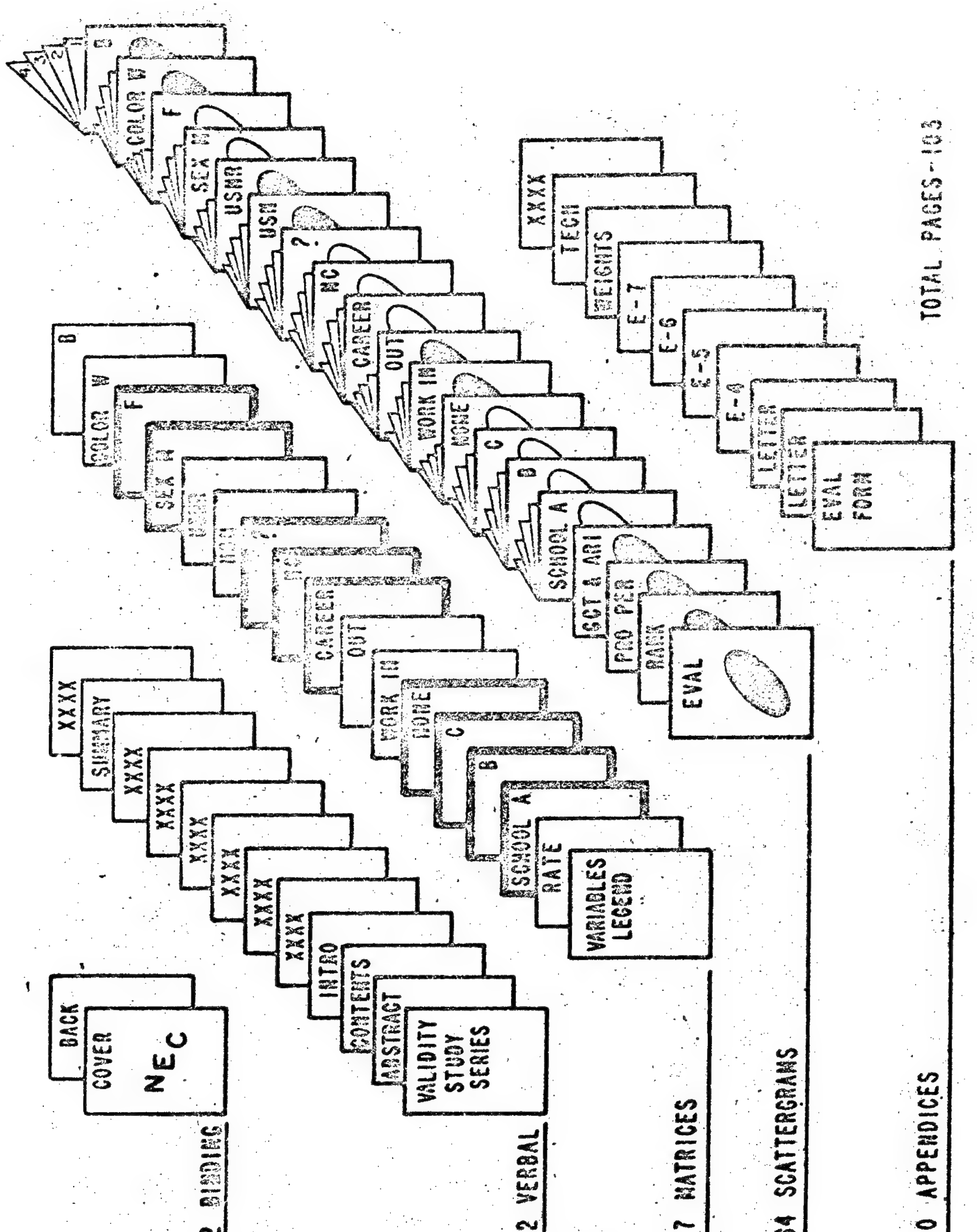

2 BINDING
12 VERBAL
17 MATRICES
64 SCATTERGRAMS
10 APPENDICES
TOTAL PAGES-103

TABLE 8

POPULATION- YN2 N= 645

28 X 28 CORRELATION MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	
1		36	82	53	44	33	35	41	36	25	-19	-20	-6	31	25	20	21	8	5	23	0	7	-9	8	-5	11	26	22	1
2			58	21	19	20	14	16	20	15	-10	-13	7	19	15	11	9	6	5	11	-3	-0	-7	4	1	1	10	11	2
3				48	39	37	31	38	38	28	-21	-20	-5	35	29	22	23	10	7	25	-0	8	-12	6	-3	9	27	24	3
4					72	51	53	68	51	23	-21	-21	-13	34	23	16	20	8	10	20	6	-4	-7	11	-3	6	22	20	4
5						39	65	75	53	16	-19	-20	-5	30	16	10	12	6	10	13	4	-3	-11	11	-3	5	22	10	5
6							40	40	38	12	-12	-13	-17	21	12	9	9	7	3	10	2	1	-7	2	4	0	17	14	6
7								57	49	14	-11	-12	-4	27	13	2	5	6	13	4	9	-2	-3	3	-6	4	15	9	7
8									50	16	-16	-16	-12	29	16	8	12	7	4	11	3	-3	-7	9	-2	5	18	12	8
9										25	-22	-27	-4	56	25	16	17	7	7	18	5	3	-14	14	-5	2	24	22	9
10											-28	-25	-18	89	100	39	39	13	7	43	10	14	-9	9	8	9	37	72	10
11												66	49	-12	-28	-51	-44	-30	-16	-53	-12	1	41	20	-9	-4	-40	-29	11
12													34	-12	-25	-26	-25	-12	-9	-29	-8	-1	31	-19	-11	1	-18	-27	12
13														-0	-17	-21	-18	-12	-4	-22	-11	-1	23	0	-7	1	-13	-18	13
14															89	30	32	10	6	34	9	13	-4	15	2	8	34	65	14
15																39	39	13	7	43	10	14	-9	9	8	9	37	71	15
16																	61	21	20	90	13	-2	-25	-13	1	15	58	27	16
17																		33	18	89	12	-5	-20	-23	10	22	47	28	17
18																			7	30	9	-2	-14	-28	-2	4	22	8	18
19																				21	-5	4	-3	-8	-1	9	10	11	19
20																					14	-4	-25	-20	6	21	58	31	20
21																						2	-2	6	-1	4	9	8	21
22																							6	10	6	5	6	16	22
23																								-3	-3	-5	-20	-5	23
24																									-4	-4	-0	5	24
25																										3	10	7	25
26																											13	7	26
27																												22	27
28																													28

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
MEAN	8.36	9.32	119.42	8.94	8.94	8.14	8.88	9.09	42.16	50.88	2.65	2.65	.31	98.46
SD	1.67	1.08	19.96	1.19	1.05	2.04	1.07	.93	4.48	10.17	1.25	1.19	.88	11.94
N	645	638	638	645	645	407	645	645	637	640	646	645	645	640

	15	16	17	18	19	20	21	22	23	24	25	26	27	28
MEAN	80.97	60.64	59.15	56.75	50.01	119.79	1.61	1.94	1.42	1.91	1.98	3.97	7.98	2.45
SD	17.84	7.57	7.39	8.33	7.24	13.43	.57	.24	.66	.28	.14	.24	2.95	.89
N	640	573	573	573	573	573	645	645	645	645	645	644	486	643

RATE- YN2
N - 364
SUBPOPULATION- SERVICE SCHOOL A
28 X 28 CORRELATION MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	
1		32	81	56	47	36	33	45	38	28	-16	-11	3	36	28	23	21	7	-0	25	0	14	-7	11	-10	15	31	23	1
2			56	19	14	10	2	12	14	17	-6	-10	10	19	17	12	7	0	4	11	0	-2	-9	7	1	6	13	12	2
3				50	40	35	27	42	37	31	-15	-13	3	39	32	19	19	2	2	22	0	13	-9	10	-6	13	31	26	3
4					76	55	56	70	52	27	-19	-16	-6	40	27	19	16	4	0	20	0	-4	-6	13	-7	9	26	21	4
5						46	64	76	51	25	-22	-20	-1	37	25	17	14	5	5	17	0	-3	-9	14	-3	6	31	16	5
6							41	46	40	19	-7	-7	-10	30	19	2	9	5	-11	6	0	-5	-7	6	8	1	17	16	6
7								59	49	24	-13	-9	-3	38	24	5	5	3	9	5	0	-2	-0	5	-6	5	25	13	7
8									51	27	-17	-20	-10	39	27	11	9	5	-2	11	0	-1	-4	14	-5	5	27	21	8
9										32	-19	-25	-4	61	31	14	11	8	1	14	0	-2	-9	17	-5	3	24	27	9
10											-31	-30	-16	93	100	38	34	9	1	41	0	15	-7	11	6	11	39	71	10
11												61	39	-18	-31	-44	-34	-26	-7	-44	0	-0	33	23	4	-6	-40	-32	11
12													22	-20	-29	-18	-17	-12	-0	-20	0	-4	31	-23	2	-3	-13	-34	12
13														-4	-15	-5	-4	-5	4	-6	0	0	11	5	2	-1	1	-15	13
14															92	33	29	7	1	35	0	12	-3	16	4	10	39	65	14
15																38	35	10	1	42	0	15	-7	11	6	12	39	70	15
16																	55	17	10	88	0	-2	-17	-9	-3	15	58	23	16
17																		31	13	88	0	-6	-10	-20	13	31	45	21	17
18																			3	27	0	-0	-14	-25	-10	13	11	5	18
19																				13	0	1	0	-2	-6	20	4	4	19
20																					0	-4	-16	-16	6	26	58	25	20
21																						0	0	0	0	0	0	0	21
22																							4	19	4	9	5	11	22
23																								-9	-2	-4	-17	-4	23
24																									-4	-4	6	4	24
25																										8	11	5	25
26																											14	10	26
27																												21	27
28																													28

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
MEAN	8.32	9.31	118.85	8.95	8.93	8.22	8.93	9.09	42.20	51.94	2.51	2.59	.23	99.47
SD	1.71	1.07	19.64	1.25	1.10	1.90	1.06	.93	4.45	10.25	.98	1.01	.75	11.81
N	364	359	359	364	364	215	364	364	360	360	364	364	364	360

	15	16	17	18	19	20	21	22	23	24	25	26	27	28
MEAN	82.82	61.24	59.87	57.46	49.67	121.11	2.00	1.94	1.41	1.92	1.98	3.97	8.29	2.53
SD	17.97	7.02	6.83	7.82	7.09	12.20	0	.24	.65	.27	.19	.18	2.91	.85
N	360	320	320	320	320	320	364	364	364	364	364	364	278	364

TABLE 10

RATE- YNP
N - 272
SUBPOPULATION- SERVICE SCHOOL NONE
28 X 28 CORRELATION MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	
1		42	63	50	38	31	38	36	35	21	-25	-29	-17	25	21	17	21	10	11	21	0	-4	-14	5	5	8	21	21	1
2			62	26	25	29	29	21	29	11	-16	-16	2	18	12	11	14	12	6	14	0	3	-6	2	1	-2	9	7	2
3				46	37	40	38	34	42	25	-28	-27	-14	31	25	25	29	18	13	29	0	3	-15	1	2	6	29	22	3
4					65	51	49	65	49	18	-26	-25	-21	26	18	12	24	11	22	20	0	-5	-9	9	4	4	19	19	4
5						33	65	72	56	1	-19	-22	-8	19	1	1	10	6	17	6	0	-5	-13	7	-4	4	12	1	5
6							38	34	34	2	-15	-20	-24	9	2	13	8	6	13	12	0	7	-5	-1	-3	-0	14	9	6
7								55	47	-3	-8	-14	-3	12	-3	-4	3	8	18	-1	0	-2	-4	0	-5	3	3	1	7
8									47	0	-15	-13	-15	15	-0	3	14	7	11	9	0	-6	-11	3	1	5	8	2	8
9										16	-25	-29	-4	48	16	16	22	4	13	21	0	9	-19	9	-5	2	24	16	9
10											-25	-21	-18	83	100	38	42	14	15	44	0	13	-10	7	11	6	32	72	10
11												69	54	-5	-25	-56	-51	-32	-26	-59	0	3	49	21	-23	-2	-43	-27	11
12													41	-4	-21	-32	-30	-11	-17	-35	0	2	32	-14	-20	3	-21	-22	12
13														5	-18	-32	-28	-16	-13	-33	0	-1	34	-2	-19	4	-24	-20	13
14															83	24	33	9	13	31	0	14	-3	13	-1	6	25	63	14
15																36	42	14	15	44	0	13	-9	7	11	7	32	72	15
16																	66	23	34	91	0	-2	-32	-19	7	15	57	29	16
17																		34	25	91	0	-5	-29	-20	7	16	47	34	17
18																			13	31	0	-3	-14	-32	9	-2	32	8	18
19																				32	0	8	-6	-14	6	2	20	18	19
20																					0	-4	-34	-26	8	17	57	35	20
21																						0	0	0	0	0	0	0	21
22																							10	1	9	1	6	23	22
23																								3	-6	-5	-23	-6	23
24																									-4	-5	-8	5	24
25																										-2	10	11	25
26																											12	5	26
27																												21	27
28																													28

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
MEAN	8.40	9.35	119.96	8.90	8.93	8.07	8.80	9.08	42.04	49.48	2.84	2.74	.43	97.06
SD	1.62	1.07	20.26	1.12	.99	2.16	1.09	.95	4.51	9.89	1.52	1.40	1.02	11.94
N	272	270	270	272	272	185	272	272	268	271	272	272	272	271

	15	16	17	18	19	20	21	22	23	24	25	26	27	28
MEAN	78.53	59.70	56.12	55.79	50.49	117.62	1.00	1.94	1.44	1.89	1.99	3.96	7.57	2.35
SD	17.35	8.16	8.00	8.95	7.47	14.70	0	.24	.67	.31	.18	.31	2.97	.94
N	271	245	245	245	245	245	272	272	272	272	272	271	203	270

TABLE 11

RATE- YN2
POPULATION- SERVICE SCHOOL
FACTOR- GCT+ARI
N- 316

FACTOR										
SS SCORE	1	2	3	4	5	6	7	8	9	10
76-80								2	2	
71-75							1	1	5	
66-70							5	7	8	3
61-65						1	2	13	11	1
56-60				1	1	5	8	21	14	2
51-55					2	9	19	18	10	2
46-50				1	3	6	20	12	7	2
41-45			1	1	1	8	16	17	4	1
36-40			1		1	4	12	5	2	
31-35			1	1	1	2	3	3	1	
26-30					1		2	2		
21-25										

CORRELATION= .4100

FACTOR(X)
STD-DEV = 12.188
MEAN = 121.197

SS(Y)
STD-DEV = 10.267
MEAN = 51.797

Y= .3461 X + (9.8632)

X= .4858 Y + (96.0362)

GIVEN X = 105.5
Y = 46.37

GIVEN Y = 48.0
X = 119.33

TABLE 12

RATE- YN2
POPULATION- SERVICE SCHOOL NONE
FACTOR- GCT+ARI
N- 244

FACTOR											CORRELATION- .4400	
SS SCORE	1	2	3	4	5	6	7	8	9	10	FACTOR{X}	SS{Y}
											STD-DEV = 14.652	STD-DEV = 9.940
76-80									1		MEAN = 117.922	MEAN = 49.525
71-75							1	1	6			
66-70								3	5	2	Y= .2985 X + { 14.3284}	
61-65					1	1	4	4	7		X= .6486 Y + { 85.7995}	
56-60							8	9	7		GIVEN X = 109.5 Y = 49.82	
51-55				1	2	7	8	13	12	1	GIVEN Y = 48.0 X = 116.43	
46-50			1	1	3	5	16	17	4	1		
41-45		1		7	6	11	10	14	3	2		
36-40			1	1	1	7	4	6		1		
31-35					4	6	2	2				
26-30						1	1	1				
21-25												

TABLE 13

RATE- YN2
N - 606
SUBPOPULATION- WORKING I/O OF RATE IN
28 X 28 CORRELATION MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	
1		38	82	56	46	33	35	44	38	26	-18	-20	-6	33	26	21	22	8	3	24	2	0	-11	11	-2	11	27	23	1
2			59	22	20	21	14	17	21	14	-10	-11	6	19	14	10	8	7	4	10	-4	0	-8	6	3	2	9	9	2
3				49	40	36	31	40	39	28	-20	-19	-5	35	28	22	24	10	6	26	1	0	-12	6	-1	9	29	23	3
4					73	52	54	69	52	26	-22	-21	-13	36	26	18	22	7	11	22	6	0	-7	13	-2	8	23	22	4
5						40	65	74	54	17	-20	-21	-4	31	17	11	13	6	11	14	5	0	-10	12	-3	6	24	12	5
6							40	40	37	14	-12	-14	-17	22	14	12	12	6	7	13	1	0	-8	0	6	2	17	15	6
7								57	50	15	-10	-13	-3	28	15	1	5	6	13	3	9	0	-2	5	-5	5	16	9	7
8									51	18	-17	-16	-12	30	17	9	13	7	5	12	4	0	-7	10	-3	6	20	14	8
9										25	-22	-27	-4	56	25	17	17	8	8	19	4	0	-15	15	-5	3	23	21	9
10											-28	-25	-18	88	100	40	39	14	5	44	10	0	-10	10	7	7	38	71	10
11												66	49	-12	-28	-51	-44	-30	-16	-53	-13	0	42	19	-10	-3	-41	-30	11
12													-35	-11	-25	-26	-24	-12	-9	-28	-8	0	31	-20	-14	2	-18	-27	12
13														-0	-18	-21	-15	-12	-5	-22	-11	0	23	3	-9	2	-12	-19	13
14															88	31	33	11	5	35	9	0	-5	16	1	7	34	63	14
15																40	40	15	5	44	10	0	-10	10	7	8	38	70	15
16																	61	21	19	90	13	0	-26	-10	-1	14	58	28	16
17																		34	17	89	12	0	-20	-20	11	20	48	30	17
18																			7	31	10	0	-15	-27	-1	3	24	9	18
19																				20	-6	0	-2	-5	-3	8	10	9	19
20																					14	0	-26	-17	6	19	59	32	20
21																						0	-3	8	-2	4	8	6	21
22																							0	0	0	0	0	0	22
23																								-4	-5	-4	-21	-7	23
24																									-4	-4	-2	5	24
25																										4	10	7	25
26																											12	7	26
27																												22	27
28																													28

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
MEAN	8.39	9.32	119.84	8.93	8.93	8.15	8.88	9.08	42.20	51.25	2.68	2.64	.31	98.85
SD	1.62	1.08	19.86	1.21	1.07	2.03	1.09	.95	4.52	10.11	1.27	1.21	.08	11.93
N	606	599	599	606	606	383	606	606	599	602	606	606	606	602

	15	16	17	18	19	20	21	22	23	24	25	26	27	28
MEAN	81.60	60.61	59.05	56.72	50.08	119.66	1.61	2.00	1.43	1.92	1.98	3.97	8.02	2.49
SD	17.75	7.62	7.37	8.39	7.20	13.43	.57	0	.67	.27	.13	.24	2.97	.87
N	602	538	538	538	538	538	606	606	606	606	606	605	460	605

TABLE 14

RATE- YN2
N - 39
SUBPOPULATION- WORKING T/O OF RATE OUT
28 X 28 CORRELATION MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1		15	83	38	22	35	41	20	23	6	-46	-11	-12	9
2			52	-6	-17	12	6	-20	14	27	-32	-40	17	25
3				34	27	43	37	19	33	18	-54	-37	-4	21
4					37	45	31	42	28	-12	-13	4	-22	-2
5						32	46	76	35	-7	1	-17	-18	4
6							27	39	52	-12	-1	11	-13	10
7								53	25	-4	-32	5	-28	2
8									29	-12	7	-3	-15	-1
9										24	-13	-2	-14	57
10											-38	-35	-3	92
11												41	34	-25
12													9	-20
13														3

	15	16	17	18	19	20	21	22	23	24	25	26	27	28	
1	6	17	12	18	14	16	-27	0	2	-15	-21	3	7	3	1
2	27	32	39	-8	10	30	9	0	4	-16	-36	-10	14	31	2
3	18	22	26	19	7	26	-20	0	-11	-11	-22	-1	21	14	3
4	-13	-22	-28	15	-10	-27	1	0	6	0	-11	-21	22	-2	4
5	-8	-10	-10	-12	-11	-11	-4	0	-13	13	2	-21	-6	-9	5
6	-12	-21	-29	14	-43	-26	25	0	19	19	-6	-19	22	7	6
7	-4	16	8	11	8	13	7	0	-11	-20	-16	-14	-7	-1	7
8	-13	-15	-15	-3	-20	-16	-12	0	-0	6	7	-21	-32	-5	8
9	24	-2	14	-2	-19	7	26	0	-0	-1	-12	-13	33	37	9
10	100	33	43	-10	26	42	7	0	-11	-12	8	18	16	80	10
11	-37	-51	-55	-33	-21	-57	-9	0	21	47	8	-30	-34	-40	11
12	-34	-34	-53	5	-15	-47	8	0	16	-8	12	-22	-25	-40	12
13	-2	-19	-15	-0	8	-19	-17	0	22	-29	11	-13	-32	1	13
14	92	22	35	-11	15	31	15	0	-4	-10	5	6	20	79	14
15		33	43	-9	26	41	6	0	-11	-11	8	18	16	79	15
16			71	11	41	92	10	0	-9	-38	19	39	35	12	16
17				20	28	93	11	0	-9	-47	4	45	21	26	17
18					4	17	4	0	-8	-39	-12	29	-25	-6	18
19						37	8	0	-18	-47	11	23	6	24	19
20							11	0	-9	-46	12	45	29	21	20
21								0	16	-19	3	-6	16	34	21
22									0	0	0	0	0	0	22
23										1	14	-27	-13	6	23
24											-12	-15	16	-18	24
25												-7	17	-2	25
26													25	7	26
27														10	27
28															28

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
MEAN	7.92	9.33	112.87	9.13	9.08	8.08	8.95	9.21	41.62	45.03	2.58	2.70	.33	92.30
SD	2.32	.94	20.40	.76	.80	2.27	.78	.65	3.92	9.26	.82	.84	.73	10.39
N	39	39	39	39	39	24	39	39	38	38	39	39	39	38

	15	16	17	18	19	20	21	22	23	24	25	26	27	28
MEAN	70.95	61.20	60.66	57.26	48.94	121.86	1.56	1.00	1.26	1.79	1.93	3.92	7.29	1.89
SD	16.10	6.74	7.55	7.43	7.83	13.23	.50	0	.44	.40	.22	.27	2.58	.99
N	38	35	35	35	35	35	39	39	39	39	39	39	28	38

THE PRAGMATIC APPROACH TO ITEM ANALYSIS

Herman A. Mahnen
United States Army Enlisted Evaluation Center
Fort Benjamin Harrison, Indiana

This paper, entitled "The Pragmatic Approach to Item Analysis" is based, in the main, upon a cookbook procedure which I developed several months ago for comprehensive item analyses of military occupation specialty (MOS) evaluation tests (ETs). The thesis of this paper is that a truly pragmatic approach to item analysis is that methodology which utilizes an optimal combination of exacting psychometric techniques, subject-matter expert opinion, and the best of the item writer's art.

The item analysis approach espoused in this paper was designed to be applicable to all MOS evaluation tests. The express purpose of this prototype item analysis format was to provide useful information and structured feedback to United States Army Enlisted Evaluation Center (USAEEC) personnel psychologists so as to facilitate the development of "better," i.e., more reliable and valid MOS evaluation tests. This paper highlights some of the more pertinent findings of the study and discusses the rationale behind the item analysis procedures utilized in the study.

All MOS evaluation tests developed at the Enlisted Evaluation Center (EEC) consist of 125 multiple-choice items which are divided, on the basis of expert opinion, into major areas (MAs). The number of MAs for any given ET may range from four to nine. Each MA is selected to represent a specific and significant facet of the MOS skill prism. Since ET items are selected from a number of subject-matter areas and then assigned to MAs, it follows

that the variable being measured, i.e., job proficiency, is actually a synthesis of various skills and job information. Therefore, MOS ETs developed at the EEC can be considered to be composite or stratified parallel tests (Tyron, 1957).

Item analysis is primarily concerned with the problem of selecting items for a test so that the resulting test will have certain specified characteristics. Going one step further, item analysis procedures should seek to determine a functional relationship between the parameters of the total test and appropriately selected item parameters, since both the reliability and validity of a test ultimately depend upon the characteristics of the items which constitute the test. It is probable that any test can be improved through the proper selection, substitution, and revision of items.

A major goal of item analysis is to obtain objective information concerning the items contained in the test. Guilford (1954) notes that this information may be utilized in several ways:

> It provides the opportunity to check up on the test writer's subjective judgment in selecting the items to compose the test. No matter how expert the item writer or the item critic or editor, such checks are still desirable, and the expert would be the first to welcome them. By experience with such checking, the test writer learns to improve in his art. He learns how examinees react to items in general and to the items of each test in particular. In multiple-choice tests he learns which distracters (wrong answers) or misleads are not functioning, as shown by their relative unpopularity. He gains new insights into the kind of item that does best in this kind of test and thinks of new hypotheses concerning the nature of the ability being measured. He learns where and how items need to be rewritten (p. 459).

The present study is mainly concerned with two central item characteristics: these are item difficulty value (p) and item discrimination value (r). Item difficulty, or the item difficulty index, is simply the proportion, p, of the tested examinees who pass the item. Item difficulty levels have been found to influence the shape of total score distractions (Ray, Hundeby, and Goldstein, 1962). Test skewness and kurtosis are direct functions of item difficulty. If a test is too difficult, the distribution will be positively skewed. If the test is of moderate difficulty for the group, a symmetrical distribution will result. Finally, if the test is too easy for the group examined, the test will be negatively skewed.

It has been demonstrated by Thorndike (1949) that an individual item exerts its maximum discriminative power at the 50%, or .5, difficulty level. The statistic $_{c}p$, which includes a correction for the influence of chance success is nearer the correct index of difficulty than is the uncorrected p. Because the evaluation tests developed at the EEC are formulated to produce a mean value of 78.1, it is necessary to strive for an average item difficulty value of .625 ($_{c}p = .625$) in order to achieve this parameter. Most of the item difficulty indices, however, should fall within a range of .25 to .85.

The p value is determined by both the intrinsic difficulty of the item and the effect of guessing. Guessing tends to make p values higher, the amount of increase being inversely related to the number of alternative responses for each item. Guessing not only tends to raise p value, but also introduces measurement error. Since the less guessing there is, the less measurement error there is, easy items tend to have less measurement error than more difficult items. Consequently the most discriminating item

tends to be somewhere between a corrected p value of .5 and 1.0. This is, however, no certainty as to what the exact value should be. One can only generate a model to predict the ideal level and then test how well the model works in practice. Employing one such model, Lord (1952) deduced that the most discriminating two-choice item would have an uncorrected p value of .85, a three-choice item .77, a four-choice item .74, and a five-choice item .69. However, there has not been enough research to determine whether those deductions, or deductions from other models, hold in the general case. At best the p values can only indicate the types of items that are not highly restricted in their possible correlations with total scores. It is far more reasonable to construct tests primarily in terms of the actual point-biserial correlations of items with total scores.

When determining the relationship between an item and a criterion, the criterion may be either an independent, external measure or the total score on a test or subtest. Correlations with an external criterion are usually considered to be indices of item validity, while correlations with the total test score (or subtest score) are more precisely described as indices of internal consistency. Empirical item validity and item internal consistency should never be regarded as interchangeable. This study addresses itself, mainly, to the internal consistency aspect of item usefulness.

The relation of validity to item statistics is more complex than for reliability. The validity of a composite of item scores depends upon both the correlation of the items with the criterion and the item intercorrelations. A cardinal psychometric principle is that the greater the item-criterion correlations and the lower the item intercorrelations, the greater the

validity of the total score. The optimal validity of a total score will be attained with different weighting for each item, in accordance with multiple-correlation principles. When items are weighted equally, as they are in EEC-generated tests, validity will be something less than optimal.

A primary objective in item analysis is the determination of the degree to which items can discriminate among individuals according to some criterion. Since an external criterion is lacking at the time these item analyses are conducted, the criterion is the score on the various subtests (major areas) of the MOS evaluation tests. Gulliksen (1950) has recommended the use of the point-biserial correlation as the index of item internal consistency. The point-biserial correlation is computed with dichotomous data on one variable and continuous data on the other (the criterion variable). Items which correlate well with total subtest scores should be retained as "good" items, whereas those with lower correlations should be eliminated. Logically, no test can be any better than the sum total of the items of which it is constructed.

Items which have low correlations with the total subtest score should be rejected in order to purify or homogenize the subtest. Thus, items with the highest average intercorrelations will be retained. Anastasi (1954) points out that this method of selecting items will increase test validity only when the original pool of items measures a single trait or skill and when this trait is present in the criterion. Most tests measure a combination of skills which are involved in a complex criterion.

.rifying the test in such a case may reduce its criterion coverage and thus lower validity. Internal consistency would appear to be a necessary but not a sufficient condition for test purity or homogeneity.

Guilford (1954) has pointed out that it is more important to analyze aptitude tests than achievement tests. This is because it is sometimes more important to have achievement test items approved by a subject-matter expert than to know their correlation with total score. At any rate, the reliability of the total score for either kind of test should be known. If the score reliability is low, there is more heterogeneity and no one item will normally correlate very high with it. A factor analysis or some other form of dimension analysis is probably a better technique to apply, to determine whether there are clusters of more internally homogeneous items to serve as new criteria for item analysis.

The population utilized in this study consisted of 208 enlisted men who took the evaluation test during February 1970 for MOS Code 91E40, Dental Specialist. The EEC-generated Point-Biserial Correlation Report was the principal source document employed in conducting the analysis. To a somewhat lesser extent, the MOS evaluation test proper was used in finalizing the analysis. Item statistics as well as total test parameters were examined and subjected to intensive statistical analysis. Particular emphasis was placed on analysis of individual item difficulty indices, mean p values, individual item discrimination indices, and mean r values since this is the basic information contained in the Point-Biserial Correlation Report.

Gulliksen's recommendations (Gulliksen, 1950) for developing and establishing procedures of item analysis were followed throughout this study. These include: (1) to establish the relationship between certain item parameters and the parameters of the total test, (2) to consider the problem of obtaining the item parameters in such a way that they will, if possible, not change with changes in the ability level of the validating group, and (3) to consider the most efficient methods, from both a mathematical and computational viewpoint, of estimating these parameters for the item.

The development of test items and the subsequent administration thereof does not herald the final step in the test construction process. Invariably, some items will be found to be ambiguous; others will be either too easy or too hard, etc. When analyzing a final form of a test, a specification of the nature of the analysis should be made. Thus, the present analysis utilized the p value or proportion answering the item correctly as the item difficulty level. Similarly, the point-biserial correlation of the item with total subtest score was employed in determining the internal consistency of the item. Nunnally (1967) states that the most important type of item analysis of achievement tests is accomplished by correlating each item with the total test score.

When items post a positive correlation with one another, this denotes that those with the highest average correlation are the best items. Considering that the average correlations of items with one another show a strong relationship with total subtest scores, the items that correlate

most highly with total subtest scores are the best items. Compared to items with relatively low correlations with subtest scores, those that have higher correlations with subtest scores have more variance relating to the common factor among the items. Hence, they contribute more to the test reliability.

Discussion:

The test evaluated in this report, i.e., 91E40 Dental Specialist, exhibits total test and item parameters that are well within EEC psychometric standards. The results suggest that extensive research prior to test construction is rewarded with highly reliable content-valid tests. Achievement tests are, of course, measures that require content validity. Nunnally (1967, p. 80) flatly states that, "All achievement tests require content validity ... such as a comprehensive measure of the extent to which men had performed well in a school for electronic technicians in the Armed Forces." USAEEC programs, such as Content Analysis for Validity Evaluation (CAVE) in which enlisted personnel evaluations on the criticality of items are used to reflect the degree of content validity within a test, are useful in this regard. Perhaps the most efficient and certain procedure for insuring content validity in a test is through the plan and procedures of construction.

Table 1

Action Codes Assigned to Test Items

Item Action Code			Number Cases
OK	=	Good item; reuse if current.	50
MA	=	Minimally acceptable item; reuse if current.	13
RV-p	=	Item at least minimally acceptable, but designated distractors are not functioning properly in reference to p-value (proportion of EM choosing designated distractor).	53
RV-r	=	Item at least minimally acceptable, but designated distractors are not functioning in reference to their correlation with their respective major areas.	3
QI	=	Questionable item; no statistical basis for judgment.	5
XX	=	Item unsatisfactory and should be discarded.	1
Omit	=	Item not included in scoring.	0
		Total	125

In Table 1 are shown the item analysis codes employed at the EEC. The item analysis indicated that some 62 items did not exhibit fully satisfactory item parameters. Of these 62 items, a total of 53 items were coded as RV-p. This denotes that the p values of these items are either too high or too low (either for the correct alternative or for the distractors). It is recommended that the distractors of these items be revised or replaced by more plausible distractors in order to lower or raise the proportion who select the malfunctioning distractor(s). If

plausible distractors cannot be developed, a new item which measures the same factor should be substituted.

There were three items classified as RV-r which signifies that the distractors of these items require revision since they do not discriminate in the desired direction. When an r_{ima} value for a distractor is too high, this indicates that an unacceptably large number of high scorers selected this alternative as the correct answer. It may be due to the item stem being ambiguous. It may also be that the distractor(s) is so close to the correct answer that the best qualified examinees are selecting the wrong alternative(s) as the correct answer.

There were five items designated as QI items. Although there are a number of reasons for an item being classified as such, most of them are ascribable to illogical or ambiguous distractors. Less frequently, they may be due to other causes such as unusual duty position items and purely academic items. At any rate, many QI items can be improved through appropriate revisions of the indicated malfunctioning distractors.

There was only one item coded XX. This item should be relatively easy to improve since it is apparent that two of the distractors were not plausible (since no one selected these alternatives), while the remaining distractor was selected by only .05 of those selected. It was recommended that all of the distractors be made more plausible. Of course, it may be that the item was simply too easy; it had almost no discriminative value. A study by Groscost (1966) indicated that XX items all improved when reused in evaluation tests.

Test construction and development should be structured toward ascertaining the extent of the relationship between formal training, OJT, and/or other training and subsequent performance of the tasks assigned in the field. If the schools which provide formal training are apprised as to how test items are functioning, appropriate modifications can be made in the curriculum to compensate for any observed deficiencies. Moreover, when certain MAs are found to produce proportionately more internally consistent items than other MAs, then increased weighting of test items in the former might be considered. Those items which show the higher or lower scoring personnel may be used as inspection guidelines for examining the on-site performance behavior of selected individuals. Certainly the program of instruction which does not distinguish the most salient points of behavioral modification by cross checks with signal characteristics of item analysis rejects objective evidence that may be applied in making the training more job relevant. While test major area weighting by the number of items inserted for evaluation is a difficult and risky process, there is some additional justification for concentrating the larger numbers of items in the major areas where consistency and discrimination are not affected by wide variance among the items.

Appendix G of the item analysis format should prove to be of great practical value to the test psychologist. This appendix is a chronological record of some of the most significant steps which the test psychologist should follow in the process of test development and/or test revision. Frequent reference to this documented checklist should insure that no step is being bypassed in the ongoing process of test development.

Adherence to and implementation of the suggestions proposed in this item analysis should result in an improved evaluation test, i.e., one which reflects desirable item and total test parameters. As Nunnally (1967, p. 244) points out, " ... regardless of what is found in item analysis, the final decision to include or reject an item is based primarily on human judgment." Thus, the combined judgment of the subject-matter expert and personnel psychologist must always play an important part in the selection and rejection of items for an achievement test. Brandt (1947) found that subject specialists with no training in the principles of measurement or in test construction produced valid achievement test items of varied types when they were oriented on test planning and item writing and supplied with appropriate materials for their guidance. Gulliksen (1950, p. 365) notes that, "The judgment of the subject matter expert must always play an important part in the selection and rejection of items for an achievement test." In the hands of the sophisticated test specialist, item analysis data can prove to be an invaluable tool of the trade. These data provide the necessary guidance for revising test items for reuse.

In summation, the comprehensive item analysis procedures presented in this paper encompassed a review of pertinent psychometric principles, a discussion of the several objectives of item analysis and a quantitative analysis of test parameters and individual item characteristics. Particular emphasis was directed at the analysis and interpretation of item p values and r values. Specific recommendations for the improvement of designated

items were made. Finally, a recommendation was made for tempering the purely quantitative findings of item analysis with the qualitative judgment and technical expertise of the subject matter specialist. This is felt to be a pragmatic approach to item analysis.

REFERENCES

Anastasi, A. *Psychological Testing*. New York: The Macmillan Co., 1954.

Brandt, Hyman. How effective are subject matter specialists in technical test construction. *American Psychologist*, 1947, 2, 311.

Groscost, J. P. *Re-used evaluation test items*. Fort Benjamin Harrison, Indiana: US Army Enlisted Evaluation Center, 1967 (Research Study #98).

Guilford, J. P. *Psychometric Methods*. (2d ed.) New York: McGraw-Hill, 1954.

Gulliksen, H. *Theory of mental tests*. New York: Wiley, 1950.

Lord, F. M. A Theory of test scores. *Psychometric Monographs*, 1952, No. 7.

Nunnally, J. C. *Psychometric theory*. New York: McGraw-Hill Book Co., 1967.

Ray, W. S., Hundleby, J. D., & Goldstein, D. A. Test skewness and kurtosis as functions of item parameters. *Psychometrika*, 1962, 27, 39-47.

Thorndike, R. L. *Personnel Selection: Test and Measurement Technique*. New York: Wiley, 1949.

Tryon, R. C. Reliability and behavior domain validity: Reformulation and historical critique. *Psychological Bulletin*, 1957, 54, 222-249.

MOS MASTERY TEST DEVELOPMENT PROCEDURE

J. E. Hohreiter
U. S. Army Enlisted Evaluation Center

A project team consisting of J. Hohreiter, F. Atchinson, R. Stitt, and A. Hermansen,was formed in January 1971 to analyze MOS test planning methods of the U. S. Army Enlisted Evaluation Center in terms of the job analysis and the testing methods of other agencies and industrial concerns. Their principal objectives were to:

1. Improve the accuracy and usefulness of evaluations of the capabilities of enlisted personnel to perform the duties of Military Occupational Specialty skill levels (MOSC).

2. Develop accurate, comprehensive descriptions of the duties, tasks, and requirements of the MOSC.

3. Evaluate the abilities of enlisted personnel to perform the full scope or range of MOSC requirements.

4. Establish minimum standards of performance.

5. Measure relative capabilities of assigned personnel.

6. Rank enlisted personnel within each skill level pay grade according to their respective performance capabilities.

The Army Military Occupational Specialty (MOS) test program was and is intended to measure the competence of examinees to perform the full range of duty positions and pay grades contained in Military Occupational Specialty skill levels (MOSC). In the majority of cases studied, the test

plans and tests developed for the program were based upon general MOS and MOSC descriptions, designed for personnel classification purposes, supplemented by recommendations from instructors for introductory MOSC training courses. Occasionally, other sources such as Tables of Organization and Equipment, Field Manuals, Technical Manuals, and text books were consulted to verify or expand data and information provided by classification guides and instructors. In some instances, Military Occupational Data Bank reports of answers of a "random" sample of personnel assigned an MOS to questionnaires concerning their current duties, or similar surveys, were used to gather additional information.

To determine whether the intent of the MOS test program was, or could be, fulfilled through application of the test plans and tests developed for the program, the project group analyzed all available information concerning three military occupational specialties. Current Tables of Organization and Equipment and guidelines for staffing Tables of Distribution and Allowances for other TOE units were reviewed to identify duty positions and roles of personnel assigned the MOS. All published operations and training publications were reviewed for descriptions of related duties, tasks, and performance standards. Units were visited. Questionnaires were sent to personnel assigned the MOSC rated above average. MODB reports and equipment inventories were analyzed. Experts were consulted. Pertinent texts and research reports were reviewed. Test plans were evaluated. The findings were enlightening but not surprising.

a. Only a portion of the requirements of the MOSC involved had been identified and considered in the construction of the test plans and tests.

b. The proportion of MOSC requirements evaluated by the tests could not be determined by examination of the test plans or of the limited background data used in the construction of the test plans. How much of an MOSC was covered by the test could be established only when the full scope of the MOSC requirements had been determined. The latitude permitted item writers coupled with the deletion of test items for purely statistical reasons further complicated the problem of identifying relationships between test coverage and MOSC requirements.

c. Opportunities to identify MOS structure strengths and weaknesses were found to be limited. Those that had been identified were usually brought to light in the course of analysis of recommendations from school instructors, complaints from examinees and commanders, and infrequent reviews of Department of the Army publications.

d. Enlisted personnel disadvantaged could not be determined from the data and materials involved in their evaluation. Test plans and tests emphasized tasks judged to be of average complexity by the item writer and test developer. Personnel assigned to duty positions beyond the scope of the test plan had no opportunity to demonstrate their extended capabilities.

e. Determinations of minimum qualification scores were found to be arbitrarily established by formulas based upon "chance" score, the distribution of examinees' scores, and the opinion of one or a small number of item writing agency representatives.

Other governmental and industrial occupational testing programs are more concerned with assessing the aptitudes of individuals for assignment to a training program or to a higher level position. Their principal objective or intention is to predict how well the individual will learn what is to be taught or perform duties to be assigned. Management oriented or employee oriented approaches are used to determine the training program or occupational requirements. The management oriented approaches involved analysis of the organization charts, position descriptions, formal training programs, manuals or directives which describe the operations personnel selected will be expected to perform, and performance or selection standards. The employee oriented approaches involve gathering of information concerning the duties, tasks, and performance standards for the positions involved from persons assigned to or supervising the positions by questionnaires or interviews. Tests are generally based upon prerequisites for the training programs or positions determined through either the management oriented or employee oriented approaches and upon statistical analyses of test scores attained by selected employees. Despite the differences in testing objectives and methods, the job information gathering and analysis steps were reviewed for applicability to the Army MOS test program.

a. Management oriented approaches can provide accurate job information more economically, if the managements directives governing employee activities are detailed and current. Requirements of vacant and planned positions can be determined. In addition, information obtained is not

contaminated with the misconceptions and imaginations of uninformed or misinformed employees. On the other hand, if the management directives are incomplete or not current, the information obtained and its applications will be correspondingly deficient and obsolete. At either extreme, the resultant job information and analyses, reflect managerial requirements for its employees.

b. Employee oriented approaches involving the use of questionnaires vary widely in quality and cost. They range from the sending of a form or letter to a small portion of the employees asking them to describe what they do - to the sending of long lists of highly structured questions concerning the details of their activities to all employees. The more economical and simpler of these approaches assume the few employees contacted can and will furnish complete, accurate, and representative information concerning the duties, tasks, and performance standards of their positions. The more elaborate systems involve applications of management approaches in the development of the questionnaires along with the additions in costs and time needed to reproduce and administer the questionnaires and process the results. None of the questionnaire approaches can be used to acquire information concerning the duties of vacant or planned positions. All of the questionnaire approaches produce descriptions of employee viewpoints of management requirements for their positions or jobs.

c. Employee approaches involving the use of interviews vary in quality and costs for the same reasons questionnaire approaches vary. Costs for

interview systems exceed the costs of equivalent questionnaire systems by the salaries and travel expense paid the trained interviewers. The potential for greater accuracy exists, however, in that the interviewer is in the position to clarify ambiguities, resolve counter-claims, and rectify omissions and oversights detected during the course of the interviews. Exploration of this additional potential increases costs and extends the duration of the study significantly. Even where resources are available for exhaustive studies, the data obtained reflects only employee viewpoints which must be further analyzed in terms of mangerial requirements.

An abridged task analysis and test planning system was developed by the project group. It is essentially a set of management oriented procedures, extended by employee oriented methods, if necessary. It leads to the construction of MOS skill level competence yardsticks based upon established MOSC requirements rather than the capabilities of individuals assigned or being trained for assignments to specific duty positions, and to test scores which describe both the examinee's relative competence with respect to others and proportionate mastery of an MOS skill level. It identifies needs for and the character of new and revised directives pertaining to MOSC requirements and related personnel management actions. It is adaptable to machine processing. It provides a base for modern mathematical analyses of test results.

The process,as outlined in the appended chart, begins with a review of authorization documents, such as Table of Organization and Equipment

and Tables of Distribution and Allowances, to identify functions and locations of MOS related duty positions established by the Department of the Army to carry out current and mobilization missions. The data obtained is compared with the MOS classification description. If the data and the MOS description are consistent, the researcher lists his findings and proceeds with the identification of the duties and tasks of the duty positions. If omissions or contradictions occur, the researcher studies operational and training publications, planning reports, and general information publications, in turn, until the inconsistency is relieved or infeasibility of the continuance of the study is established.

After the duty positions of the MOS, their organizational associations, their mission relationships, and the types of equipment persons assigned the duty positions must operate or repair has been determined, operational and training publications are studied in detail to identify mission related duties, tasks, and performance standards. If the information obtained covers the functions of all identified duty positions, the researcher develops a description of each skill level of the military occupational specialty. Omissions or contradictions discovered are reconciled to the extent possible by study of supplemental publications such as Maintenance Allocation Charts, New Equipment and Personnel Requirements Summaries, systems engineering studies, manufacturer's manuals, contract studies, and general publications. Any remaining discrepancies are relieved by personal or telephone contacts with officials of functional agencies, telephone interview surveys of unit commanders' opinions, contacts with

MOS experts or study project heads, questionnaires addressed to superior personnel assigned to duty positions involved, or on-site interviews with personnel concerned and their commanders.

The duty positions, related duties and tasks, and performance standards are then grouped according to MOS skill level classification descriptions and are classified according to the level of skill and training required for their satisfactory performance. Tasks which are,or can be, performed by individuals who lack minimum qualifications for award of the MOS skill level are assigned the symbol "0". Tasks which are ordinarily assigned to, or can be performed by, those who have completed formal, individual training programs but have not completed their initial unit training program are identified by the symbol "1". Tasks which are, or can be, indiscriminately assigned to individuals who have successfully completed their individual and unit training for the MOS skill level are annotated with the symbol "2". Tasks which are assigned to selected individuals who have significantly more intensive and extensive experience than is afforded in the unit training program and routine duties are assigned the symbol "3". Tasks which are ordinarily reserved for and assigned to the most highly qualified person available, or performed in their absence by higher level persons, are identified by the symbol "4". Note that the symbols represent categories which are inequalities and are not elements in the set of natural numbers. They are ordered categories which are neither associative, commutative, nor normally distributed.

The detailed listing of the duty positions, duties, categorized tasks,

and performance standards for each skill level is then reduced by deleting the tasks in category "O", the duties comprised exclusively of category "O" tasks, and any duty positions which involve only category "O" tasks. The abridged listing describes the skill levels of the MOS in terms of the duty positions, duties, and tasks which transcend minimum qualification requirements and provides the base for development of plans for evaluating the competence of individuals to fulfill the full range of requirements for each skill level of the MOS. The listing is forwarded to the test construction psychologist and the item writing agency for information and their recommendations for additions or deletions supported by reliable and valid evidence.

During and following coordination of the abridged descriptions of the above minimum requirements of the skill levels of the MOS, the tasks listed are analyzed with respect to testing method suitability. Tasks in which problems, identifications, and solutions are determinants of performance capability and which can be presented in multiple choice form are identified for incorporation in the written test plan. The remaining tasks, in which physical skills that necessitate observer reports of the performance of examinees are determinants, are noted for consideration in rating scale plans. Both task groupings are finally further condensed by deletion of those affected by administrative restrictions and by stratified sampling of the balance to limits determined by available funds, machine capabilities, and other resources. The condensed task groupings are regrouped to provide meaningful subscores and annotated by the researcher as necessary to guide the test construction psychologist and item writer in the construction of test and rating scale questions.

REFERENCES

CON Reg 350-100-1, Systems Engineering of Training, Headquarters, United States Continental Army Command, February 1968, pp 8-20.

Cormack, Bruce, "Job Analysis in the Canadian Armed Forces," Proceedings 10th Annual MTA Conference, Air Force Human Resources Laboratory, September 1968, pp 227-232.

Handbook for Construction of the SKT and Associated Tests, Air Force Human Research Laboratory, October 1968, pp 1-1 through 14-2.

Hohreiter, J. E., "Programmers Please Perceive," Proceedings 7th Annual MTA Conference, 6570th Personnel Research Laboratory, USAF, October 1965, pp 11-18.

Item Writer's Guide, U. S. Army Enlisted Evaluation Center, January 1970, pp 2-1 through 6-7.

Loomis, Donald O., "The USAF Operational Job Analysis Program," Proceedings 10th Annual MTA Conference, Air Force Human Resources Laboratory, September 1968, pp 233-237.

Mayo, C. C., Report AFHRL-TR-69-27, Construction and Administration of Ten Air Force Job Inventories, Air Force Human Resources Laboratory, October 1969, pp 1-23.

McKnight, A. James, Report AD 649866, The Utility of Data from Field Performance Measurement, Defense Documentation Center, Defense Supply Agency, November 1966, pp 1-6.

Meyer, Harry J., "The Military Occupational Information Data Bank Output Reports and Application," Proceedings 10th Annual MTA Conference, Air Force Human Resources Laboratory, September 1968, pp 251-273.

Naval Examining Center Instruction 1418.6A, Developmental Instructions and Procedures for the Examination Development Department, Naval Examining Center, January 1971, Enclosures 1-12.

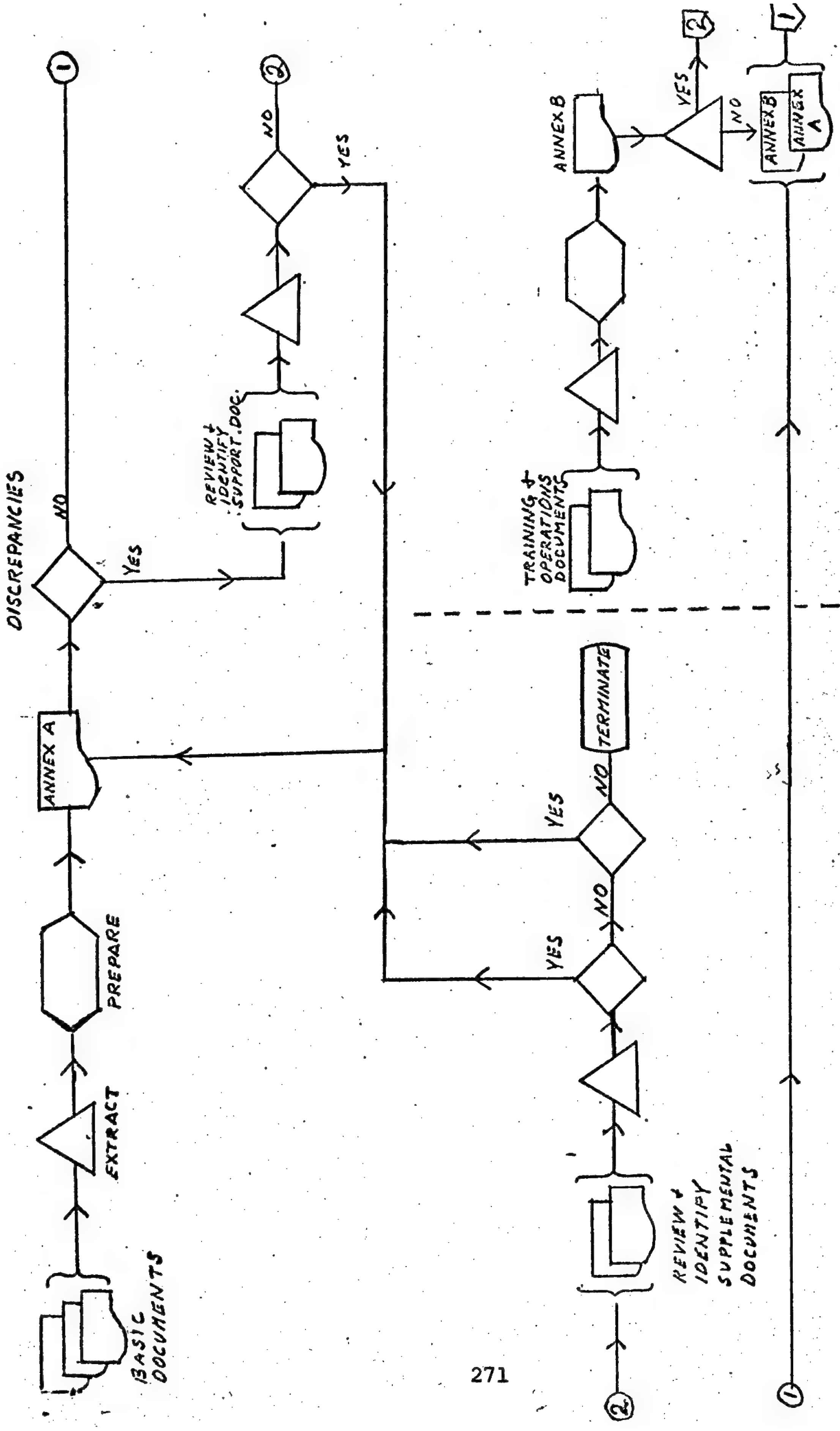
Rundquist, Edward A., Research Report SRR 71-4, Job Training Course Design and Improvement, Naval Personnel and Training Research Laboratory, San Diego, California, September 1970, pp 33-69.

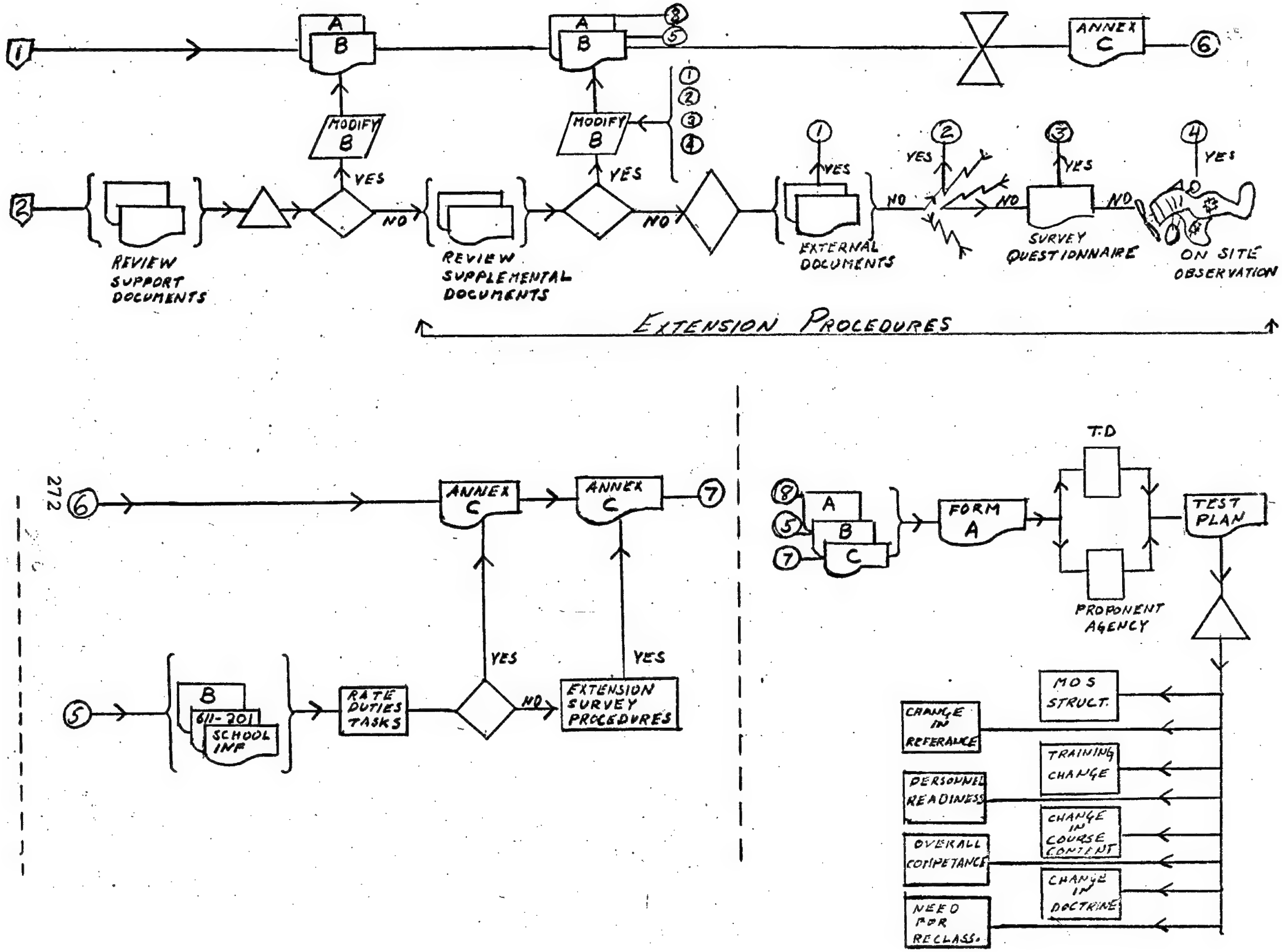
SDB Report No. 1-60-OR, Army Job Analysis Manual I, Research and Development Division, TAGO, Department of the Army, March 1960, pp 3-70.

Standing Operating Procedures for Construction of MOS Evaluation Tests, U. S. Army Enlisted Evaluation Center, June 1961, pp 1-1 through IV-7.

Tracey, W. R., et al., The Development of Instructional Systems, US Army Security Agency Training Center and School, December 1970, pp 1-1 to 2-5, 5-16, 6-9.

Wallace, S. Rains, "The Relationship of Psychological Evaluation to the Needs of the Department of Defense," Proceedings 7th Annual MTA Conference, 6570th Personnel Research Laboratory, USAF, October 1965, pp 11-18.


BASIC DOCUMENTS
EXTRACT
PREPARE
ANNEX A
DISCREPANCIES
YES
NO
1
REVIEW + IDENTIFY SUPPORT. DOC.
YES
NO
2
REVIEW + IDENTIFY SUPPLEMENTAL DOCUMENTS
YES
NO
YES
NO
TERMINATE
TRAINING + OPERATIONS DOCUMENTS
ANNEX B
YES
NO
2
ANNEX B
ANNEX A
1


A
B
MODIFY B
YES
NO
REVIEW SUPPORT DOCUMENTS
REVIEW SUPPLEMENTAL DOCUMENTS
EXTERNAL DOCUMENTS
SURVEY QUESTIONNAIRE
ON SITE OBSERVATION
ANNEX C
EXTENSION PROCEDURES
611-201
SCHOOL INF
RATE DUTIES TASKS
EXTENSION SURVEY PROCEDURES
FORM A
T.D
PROPONENT AGENCY
TEST PLAN
MOS STRUCT.
CHANGE IN REFERANCE
TRAINING CHANGE
PERSONNEL READINESS
CHANGE IN COURSE CONTENT
OVERALL COMPETANCE
CHANGE IN DOCTRINE
NEED FOR RECLASS.

"Relative" Rating System & Small Ratee Groups

A Paper Presented to the
Military Testing Association
Sept. 20-24, 1971

by

Kirt E. Duffy

Personnel Systems Branch
Personnel Research Division
AF Human Resources Laboratory
Lackland AFB, Texas

"Relative" Rating Systems and Small Ratee Groups

Rating systems which force discrimination have been suggested as alternatives to systems which are not so constrained, when the latter break down by failing to yield necessary discrimination. Systems which force discrimination, for example, rank ordering and pair comparisons, have been called "relative" as opposed to "absolute" rating systems. The intention of this paper is to show that these relative rating systems guarantee excessive error as the size of the ratee groups become small. The "point allocation technique" will serve as a paradigm, despite not being a pure example of discrimination forcing, since it has the property of being capable of unlimited discrimination (for groups greater than one) and is particularly demonstrative of the general problems as well.

In the point allocation technique the rater is given a certain number of points per ratee in his group, and then allocates these points among the ratees. This method does not guarantee discrimination but ensures a direct comparison among the ratees such that, for example, all the ratees cannot be rated highly. In order to see the shortcoming of this technique imagine that each ratee in the entire population has a "true" score, and that these scores are normally distributed with a mean of 100 and a standard deviation of 10, as represented in Fig. 1. The rater receives 100 points per man. It will be possible to give each man in a ratee group his correct "true" score only when the group mean is 100. This condition is exactly met if one rater rates the entire population. However, if the ratee group comprises less than the entire population, statistical sampling considerations ensure that the mean is likely to diverge somewhat from 100. In this case the rater will have the wrong number of points to allocate so that his ratings must be somewhat in error. The importance of the error depends on the ratio of the amount of error in the mean to the number of ratees among which error must be divided. Unfortunately, as this number of ratees, or ratee group size decreases, error increases, inversely as the square root of group size.

The effect of this error is demonstrated by comparing groups of size 100 with groups of size two. As seen from Figure 2, for 95% of groups of 100, the mean error does not exceed about 2, so that no more than 200 error points, or two per man, would have to be allocated. A corresponding figure for groups of size two, as seen from Figure 3, would be 14 points per man. This would mean that if one ratee received his "true" score in this situation, the other's score would be 28 points in error.

The sampling problem of the point allocation technique is present also in rank ordering procedures. However, the demonstration is complicated by the fact that, in addition to sampling error, possible discrimination is directly limited by group size, introducing a second type of error. Because the point allocation technique is limited only by sampling error, it is the clearest illustration of the effect of this error as ratee group size decreases.

A number of suggestions have been made with respect to the shortcomings of "relative" rating methods in small groups.

One argument is that an "absolute" rating procedure can be used in conjunction with the "relative"method. The "relative" dimension would enforce discrimination while the "absolute" dimension would allow for variation in the overall ability of a group. But since it is already known that the "relative" approach is a failure, it can't help a nonworking "absolute" system. On the other hand, if the "absolute" system works, the "relative" one won't add anything, since exactly the same arguments applied to the whole population apply equally well to the subpopulation at a given level of the absolute scale. See Fig. 4.

The superior officer is most unfavorably affected by an error since he is more likely to come out looking worse than he actually is. Therefore he is relatively worse off in a smaller group than a larger group. This leads to the concern that the superior officer should be able to receive a score at least as large as he deserves, or, in terms of the point allocation technique, to the suggestion that smaller groups receive more points per man. However, this only intensifies the problem since the mean is still fixed across groups, and the error about the mean is even greater.

It has been suggested that the group mean be fitted to the group by applying a "quality factor" to particular jobs, on the basis of the feeling that different qualities of job tend to be filled by different qualities of men. But again, even if this could be done, the same arguments applying to the population apply also to a subgroup at some quality level. See Fig. 4. What this would amount to would be rating according to job.

It would appear as though the solution to lack of discrimination in rating systems where ratee groups are small is not to be found in "relative" rating methods.

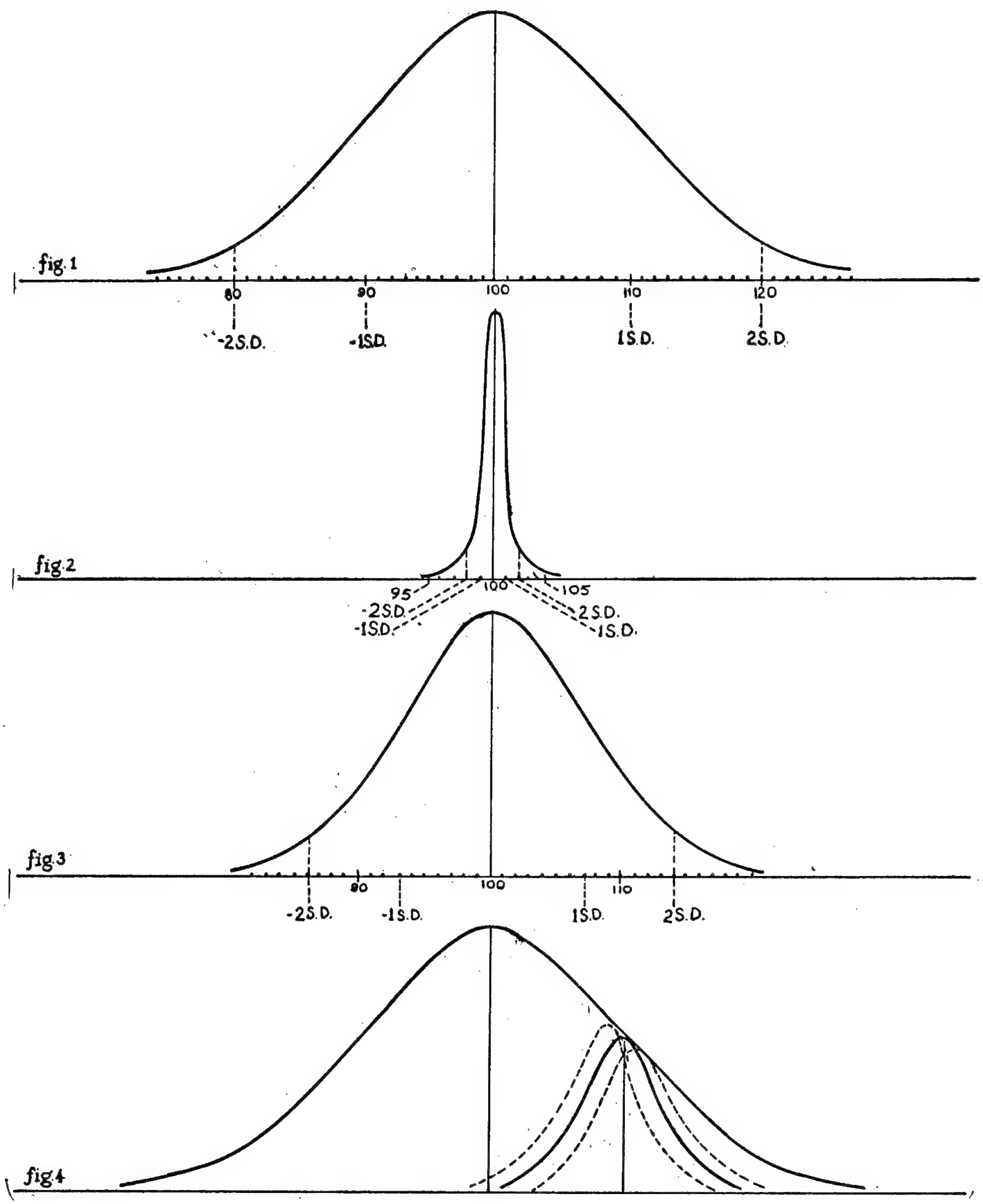


fig.1
80
90
100
110
120
-2S.D.
-1S.D.
1S.D.
2S.D.
fig.2
95
100
105
-2S.D.
-1S.D.
2S.D.
1S.D.
fig.3
90
100
110
-2S.D.
-1S.D.
1S.D.
2S.D.
fig.4

MINI-SURVEYS

Arthur G. Hermansen
United States Army Enlisted Evaluation Center
Fort Benjamin Harrison, Indiana

The United States Army Enlisted Evaluation Center (USAEEC) has the responsibility within the United States Army of developing evaluation tests (ETs) to measure job proficiency in the various military occupational specialty (MOS) codes of the enlisted MOS classification structure.

In preparing military occupational skill level tests there is a requirement for job information in order that the test development specialist can select functional job proficiency measuring questions. The job information that is required must be sound, quality, information that will produce the formation of questions that are 1) easily understood, 2) eminently fair, 3) discriminate delightfully and 4) can be crowded into a test of 100 or so multiple choice questions. In addition, the test must possess the virtues of the multiple requirements of validity and staunch reliability.

There are many methods available to the test psychologist for collecting job information. Not doubting the efficacy of prayer or the nagging correlation of extra-sensory perception, testing psychologists have traditionally turned to the survey as a means of collecting required information. It is important at this time to note that the survey <u>is</u> a proper and effective means of seeking information. The trick is to tailor your questionnaire to specific requirements.

An evaluation of some recent efforts indicates that contemporary survey methods encounter a number of serious problems. One problem appears to be that of including too many items. There appears to be a trend towards following the adage that if 100 questions are good, 200 are

better. Some questions are better left out of the survey if they really do not pertain to the major problems you are trying to answer. When we ask questions about too many different items, we tend to reopen old irritations and magnify minor problems into major proportions.

Another problem is the time lapse between survey and report of results. To be effective and to create a better climate for future survey work, the population utilized in the survey must be made aware of changes created by their response to the survey effort. A further consideration in this area is the fact that we seldom, if ever, convey our results to those who participate in the survey or even thank them for their participation. It may in some measure account for the lack of enthusiasm with which service people greet the news that they have been selected to participate in a survey.

Another limitation is the fact that items are not ranked in importance. The penalty for asking too many questions results in reporting in too many areas of consideration and clouds the issue of what answers you were really seeking. It tends to divert problem-solving energy into too many diversified channels.

Level of language is another limitation. Questions and survey results are not always presented in a format which blends into management utilization. Test psychologists too often speak in terms other psychologists will understand and do not make clear, on a management language level, just what was accomplished by the survey and how it pertains to the immediate problem.

A different approach is the mini-survey. In some respects it's like trying to duplicate the brief, informal inquiries that occur during the day when one person asks another, "How's it going?" Or the informal

unstructured "survey" that occurs after work when a group of servicemen has gathered together at a local tavern or NCO club to have a few beers. In that kind of relaxed atmosphere you often get very direct, and sometimes quite blunt comments concerning what irritates them and what problems they have on the job, what's wrong with the system, and what's so hard about the job.

For realistic test purposes we can learn from this informal experience about what is critical about the job and what is difficult about the job. Since we have so many constrictions on what we can put into a proficiency test, we would certainly want to include questions primarily directed towards criticality in a military occupational skill level and what is difficult about the MOS job requirements.

This, then, is the very essence of the mini-survey 1) simple open-ended questions asking about, What are the most critical requirements about the equipment you work with? 2) What are the most difficult things you have to do on your job or with the equipment? 3) What takes the most training or skill to do? 4) Did you receive enough training to do the job? 5) If not, what would you suggest for training? 6) How can we improve your job?

Of course, these are only samples, but they do hit on the realistic level of the worker in the world of work. Keep the survey short, two pages at the most. Keep to the open-ended question, worry about the classification of the answers when you have all the answers at hand. Use the results as quickly as possible. For example, on a quick hitting mini-survey used with an Air Defense Artillery battalion we were able to collect data for,

recommend changes in training requirements, provide the basis for possible equipment modification, and contribute to defining areas of vital significance in the proficiency testing program. Not bad for a two page survey.

Perhaps this idea is not particularly new. It is, however, timely when you want information quickly. Have you tried it? If not, perhaps you will find it useful.

TRAINING TACTICAL DECISION MAKERS

By

Robert E. Loehe

Marine Corps Development and Education Command
Quantico, Virginia

TRAINING TACTICAL DECISION MAKERS

Student evaluations in an academic environment serve many purposes. They can measure student ability or willingness to learn, effectiveness of instructors or instructional techniques, usefulness of training aids and facilities or general effectiveness of a total training system. In turn, these measures can be used to grade students, promote or discharge instructors, improve training facilities, increase or decrease training time, or direct the student toward further learning in areas in which he is deficient. This last purpose is the one to which this paper is devoted.

The training required to develop competent tactical decision makers, that is, military officers commanding troops in tactical combat, is extremely complex. All the skills required of a business manager must also be mastered by the tactical decision maker. Thus he must deliver the right goods in the right amount to the right place at the right time. His problem, however, varies from that of businessmen in two important respects. First, he usually has much less time in which to make his decisions and second, his payoff is measured in human lives rather than dollars. How does this affect his training requirements? If his training is inadequate the results can be disastrous. That much is obvious. The central training requirement imposed by these differences is, however, the one that deals with rapid decision making under stress. More often than not there is little time for detailed analysis of the problem. The decision must, in most cases, be nearly instinctive. How then is this instinct gained? T. E. Lawrence has said: "Nine-tenths of tactics are certain, and taught in books: but that irrational tenth is like the kingfisher flashing across the pool, and that is the test of generals. It can only ensue by instinct, sharpened by thought practicing the stroke so often that at the crisis it is as natural as reflex." That elusive and irrational tenth of learning is what our decision makers must achieve.

If this elusive tenth of learning was easily identifiable, then it too could be neatly documented and given its place in the books from which the other nine-tenths are learned. Our problem would then be solved. Unfortunately, this is not the case. This tenth seems to be uniquely tailored to individual commanders, and the units they command. If our perception of this increment of learning is indeed correct, then it would appear that each man must uncover it for himself by being able to practice decisionmaking under stress and observe the outcome. It is our intent to create a training environment under which the student decision maker can experiment and attain for himself this essential element of his education.

This is, of course, easier said than done. We are not, however, starting from scratch. We have attempted and been at least partially successful in teaching this important element. Outside of actual combat we have employed two basic techniques:

The tactical field exercise and the map exercise. Both of these techniques, however, are deficient. The field exercise is expensive, time-consuming and only relatively few students can be trained at any particular time. The map exercise lacks realism. It is not capable of developing the really dynamic, high stress, environments that are necessary to impart this kind of training to the decision maker.

I intend to accomplish three things in this presentation. First, I will explain why a new system for training tactical decision makers is both necessary and feasible today; I will then discuss the technical concept of such a system; and finally, I will describe the progress that has been made toward its achievement.

Let us now look at how the need for such training has intensified over the past few years. If we go back to WWII we find that officers could be given basic training for conventional warfare, then through combat experience they could gain insight into the application of this training. This experience served well for a broad base of subsequent combat operations. This was true because the preponderance of these operations, at the combat unit level, had a great number of similarities. They were largely high intensity conventional operations. Though the experience gained must have been costly in human lives it was the only course of action available in the time allowed. The skills and knowledge attained through these experiences have largely been responsible for the combat readiness we have been able to maintain in the interim between WWII and the present. These experiences served us both in Korea and Vietnam.

Vietnam has, however, given us a preview of things to come. Our earlier experiences were not always applicable to the environment found there. New command skills were required to capitalize on more sophisticated command and control systems, more accurate and powerful weapon systems and more rapid and flexible mobility systems.

The need for this new command ability was further generated by an elusive enemy who mixed with the civilian populace and thus compounded the problem.

Our future commander must now face the requirement to deploy under a spectrum of situations ranging from influence projection through low intensity counterinsurgency operations to general war. He will be allowed no delay in responding to these emergencies to practice his skills. We must, therefore, more fully exploit the amount of time he is able to devote to training now.

This clearly points to an intensified need to develop a better tactical training system. Now we must see if a better system is feasible. Two developing techniques lead us to believe such a system development is possible. These techniques are those employed for war gaming and research simulations.

Manual war gaming has served to structure and organize the problems of representing opposing forces prosecuting conflicting objectives. As in any game, it must be played with rules, and the quality of these rules determines the validity and usefulness of the game. Volumes of rules are now available which cover nearly any event or contingency which may occur in the course of a game. But manual war gaming is a lot of work, is slow and requires a huge staff to apply the rules and derive assessments from them. War gaming, particularly at the smaller unit level where high fidelity is required for any meaningful analysis to take place, is normally run on a discontinuous time basis. The training that could be derived from this method is not dynamic and cannot approach the complexity of a real combat environment.

Research simulations on the other hand have been able to add dynamics to the problem of replicating combat, but in doing so, have largely eliminated role-playing by the participants and they have been misconceived in purpose.

The models have largely been used to provide point prediction of system effectiveness, i.e., to attempt to answer the question of whether system A is better than system B and, if so, by how much?

For this evaluation, model makers have attempted to build, refine, and manipulate scientifically-based, detailed models employing Monte Carlo techniques. These models would presumably handle the interrelationships of important variables at their absolute values. But this approach is limited by the inability of the results to be validated by an adequate base of empirical evidence.

Even though volumes of literature have, through the years, been devoted to analysis of combat, there is still insufficient empirical data to provide a scientific measurement link for completely validating a combat simulation model. Of course,

the model can still be tested against judgment by experienced officers, and can thus assume a degree of confidence by their concurrence and approval, but there is no scientifically valid basis for the confidence level. Despite their combined experience, all the officers can be wrong in their assessment of a specific complex situation and military history is replete with examples. The use of this approach for point prediction is not supportable; the best that can be expected is identification of trends. Another drawback is that the number of replications required to gain statistical stability during the operation of Monte Carlo models tends to obscure the way in which variables interact. Thus, even for those variables which have been quantified accurately, the trends are not made readily apparent when the results undergo analysis. The sheer volume of data obtained from model manipulation makes parametric analysis difficult to perform.

It is because of the difficulties involved in analysis that these detailed model systems have never been satisfactorily adapted to the function of training military tactical decision makers. Sound tactical decisions require an understanding of the effects of the change in variable values on battle outcomes. This understanding is not scientific in the sense of a capability to accurately predict detailed outcomes, but rather artistic in the sense of knowing what trends are produced through variations of the recognized elements that make up the combat phenomenon. Clearly, the mix of possible elements making up this phenomenon is nearly infinite: Some can be modeled and others cannot. Certainly no one pretends to be able to handle all of them, particularly such subtle factors as motivation of human behavior. No one has ever been able to model all the important ones. For instance: When does suppressive fire suppress the enemy?

This does not, however, imply that modeling ground combat is a useless activity. But it does mean that it must be viewed as an intellectual activity vice a scientific one. In this view, the model maker seeks to gain greater insight into the effects on battle caused by changing the value of each variable throughout its range. This change in perspective makes the activity directly applicable to the education of ground combat officers.

Combat simulation that can build this kind of insight affords an improved alternative to the school problem technique of training. Instead of using sterile, narrow "approved" options that are demanded by the school problem, the officer can try innovative applications of combat power and techniques. The realistic combat simulation model will feed back reinforcement for innovative decisions that represent a sound understanding of combat.

Manual war games have been utilized to train combat commanders, but the combination of cumbersome manipulation of apparatus and boringly slow assessments and techniques lead to a lack of dynamic play. Lack of dynamic play, in turn, impedes the feedback of stimulated responses; of decision outcome; of reward or punishment. The slow play is tedious and fails to place real-time stress on the student because it allows him extravagant time for his decision. Computer simulation avoids these delays and makes the play of the game dynamic but it suffers the even worse fault of eliminating role-playing, i.e., the place of the student in the decision process. Previous attempts to introduce human decisionmaking through computer simulation have been wrecked on the shoals of point prediction. They have either only been able to address the low resolution type of problem, that is, one in which operations are not simulated in great detail, or they have introduced an overly simplistic decision role. By realizing that the decisions are artistic and intellectual rather than scientific, and by realizing that we need only demonstrate the trends of the results of decisions, we are able to marry the manual war game and the research simulation model.

TESE, the Tactical Exercise Simulator and Evaluator, utilizing sophisticated techniques but simple parametric changes will allow the student to gain insight into the combat process.

This is the path that we intend to follow in developing our training system. Thus we hope to generate a model that will allow the student to investigate his perceptions of tactics required, test them in a realistic simulated environment, and observe the trend of results. The <u>trend</u> of results is the key concept. We do not require that the model be capable of accurate point predictions of combat, but rather that the student be able to distinguish the parameters of the total problem to which the outcome is particularly sensitive. He will then be able to analyze and evaluate his own effectiveness in these sensitive areas and be led back to the classroom and his books to increase his knowledge in these areas. He can also directly evaluate his performance in given tactical situations against his peers and against historical records of probability distributions of outcome for various tactical decisions.

The requirements for the TESE system as we now see them may be grouped for convenience into two categories, philosophical and physical.

The TESE must meet several philosophical requirements to be an effective educational aid. First of all it must be "reasonably" realistic. Recall we have said that the model need not give point predictions of the outcome of combat. This is not to say that the model should not be realistic. The model is not simply a device to force students through the process of command and staff functioning. The students will learn tactical lessons from the model that they can apply in combat, and we must take great care that the trends indicated by the model do not convey or lead to faulty or suspect conclusions. And of course, the model must be accepted by the students as being a realistic representation of combat. Furthermore, the model must be fun. It must be interesting and easy to use, not requiring significant effort to learn to manipulate the hardware. Most importantly, however, the TESE must allow the student to evaluate his own performance, discover his own weaknesses, and provide positive reinforcement for actions leading to correction of those weaknesses.

From the physical viewpoint at this stage we only envision the TESE as a computer assisted two-sided war game. Of course it should make no unusual power demands and it must be capable of displaying comprehensive real-time analysis of game play to the problem director. Thus, we are only planning for TESE in terms of the model itself, leaving the necessary hardware unspecified. We believe that this approach will leave our options open, both in the purchase of presently available hardware and in taking advantage of possible breakthroughs in the state-of-the-art.

A realistic combat simulation model that can be employed academically to improve the cognitive mental pattern in the minds of tactical decision makers does not yet exist. But we are optimistic that it can be developed for TESE. We believe the best approach lies somewhere between the manual model's aggregate assessment techniques and the detailed model's stochastic techniques.

The cornerstone of the TESE concept is the use of self-evaluation to enhance the learning process. Hopefully, the student will apply the lessons he has learned from classroom lectures and the study of textual material to combat situations through the use of TESE. He will see the results of his decisions in important areas and will confirm his lessons through self-evaluation.

As the student manipulates or "plays with" the TESE model by varying tactics, mobility factors, levels of supporting arms, logistic requirements and other parameters, he will discover

what factors are important (i.e., those factors to which the model is sensitive) for different types of problems. For example, if the model rewards effective use of supporting arms, the student will be led to investigate the techniques for utilizing and calling upon available support. His studies will then be rewarded in later play by reduced casualties or reduced time to complete his mission. The TESE must be carefully structured to ensure that (as in real combat) better techniques receive a "payoff."

Evaluation of student performance is a consideration in any training program. The purpose of all training programs is the transfer of the training to the operational environment. This is easy to measure in such areas as typing or welding. When one approaches an area such as command and control, the ultimate yardstick of the real world is absent. For this reason some intermediate criteria must be used.

Intermediate performance criteria are commonly referred to as criterion -- referenced or norm-referenced. Criterion-referenced criteria are those which measure the student's performance against some absolute standard or quality. In our case this would be the school solution to the problem. At this time we do not intend to employ this evaluation device since we are striving for a flexibility that is incompatible with its use. Norm-referenced criteria, on the other hand, evaluate a student's proficiency in terms of a comparison between his performance and that of other members of some group.

The norm-referenced criteria may prove to be effective to give the students "benchmarks" by which to judge their performance and, therefore, will be utilized in TESE. A group of students running similar problems can be given results of others in the group. The student would then be able to evaluate his performance against the benchmarks and learn through experience by repeated simulation of the same situation. This type of self-evaluation in comparison to benchmarks should lead to meaningful discussions among the students. Another type of benchmark may be used by giving the group both the standard and the best results achieved in the past. The specific tactics used to achieve good results may be discussed, but care must be taken not to present these tactics as doctrine. Instead they should be used as illustrations of techniques and factors that are important in certain types of situations.

As can be seen, the use of a standard school solution as an evaluation tool will be avoided in TESE. Insofar as evaluation consolidates and enhances the learning process it will be employed. But evaluation for the sake of a grade is incompatible with the TESE concept as we now perceive it.

The TESE project is still in a very early conceptual stage. Thus only broad operational capabilities have been described in the TESE advanced development objective.

The initial goal of our development effort is to provide a combat simulation model for the Marine Amphibious Unit, a basic Marine Corps task organization that combines the elements of both air and ground combat into a single striking force. Based on an infantry battalion landing team augmented by tanks, artillery and both fixed and rotary-winged aircraft, it usually consists of about 2,200 men. The model will be used to exercise the students of the Amphibious Warfare School in tactical decision-making at this level. The ultimate purpose, still far in the future, is to expand the scope of our first model for other, higher level applications in the Marine Corps Officer Professional Education System.

TESE development will proceed in four phases. The development effort will be reviewed at the end of each phase to determine if it is feasible and desirable to proceed to the next phase. Decisions as to termination, continuation or modifications can then be made before proceeding with new phases.

PHASE I: Determination of Methods and Requirements

The simulation requirements, as now stated in the advanced development objective, must be refined and made more explicit. These are "user's" requirements and they must be related in detail to the learning objectives of the Amphibious Warfare School. It is also necessary to determine simulation methods that show promise of meeting these requirements and devise experiments that will provide an objective basis for selecting the best technique or combination of techniques. Resource requirements for conducting such experiments must be identified. For these tasks, very close liaison must be maintained with the instructors at the Amphibious Warfare School. Not only must their military experience be obtained to assure that the TESE model will satisfactorily replicate the demands of all phases of ground combat, but their educational expertise is needed to establish the role TESE will play in the overall educational process.

The output of Phase I will be a detailed plan for proceeding into the second phase of development. The present advanced development objective will be revised as necessary and a technical development plan will be prepared to guide remaining development.

PHASE II: Research and Experimentation

Phase II will not commence until the plan developed in Phase I has been fully evaluated. If judged economically and technically feasible, research will be undertaken to define total system characteristics for the Marine Amphibious Unit level model. Experiments devised in Phase I will be conducted. It is envisioned that these experiments may take the form of pilot exercises on a miniature scale with, perhaps, the Amphibious Warfare School staff acting as aggressors and students making the necessary decisions to illustrate application of the simulation method under consideration. Whatever the form of the experiments, it will be vital that they provide an objective measure of effectiveness of the various approaches. The intended outcome of this work is a set of detailed specifications for the Marine Amphibious Unit model and its use, together with a cost analysis that will delineate the resources required to meet these specifications.

PHASE III: Production and Employment

Presuming the adequacy of specifications developed in the second phase, work will proceed to construct the model and the supporting facilities that will be required to operate it at the Amphibious Warfare School. Model construction will entail development of system logic; accumulation of data; software and hardware preparation and assembly; and coding and validating the simulated activities. Physical facilities will include requirements for a TESE control center, various peripheral equipment and installation of display and communications facilities in Amphibious Warfare School classrooms.

The combat simulation model would then be employed to support the Amphibious Warfare School curriculum. Its primary use would presumably be to replace the current map exercise. It is to be expected that a wide variety of possible applications will have been considered in the preceding phases. Forms of employment that are not now apparent may, accordingly, be developed and selected.

PHASE IV: System Improvement and Expansion

The Marine Amphibious Unit model developed in Phase III is essentially a pilot model. Although it will, hopefully, provide direct and immediate educational benefits for Amphibious Warfare School students, its development purpose is to provide a capability for evaluating the simulation process for applications in both officer professional education and in Fleet Marine Force training.

Phase IV, therefore, depends upon the success of the Marine Amphibious Unit pilot model and the learning achieved through its employment. If the model works for the Amphibious Warfare School, there would appear to be no reason why variations of the model, based on similar principles, would not be useful for other applications.

There are no time limitations on this phase. Decisions on the direction and scope of expansion will emerge through actual use of the Marine Amphibious Unit model and the educational and training requirements of other potential users of the TESE system.

As has already been noted, the TESE project is still very early in the conceptual stage. The initial military requirements document, the advanced development objective, has been promulgated by the Commandant of the Marine Corps. This document specifies the operational need for a computer simulation model of ground combat which will permit students to exercise their knowledge of tactics by real-time interaction with the model.

The Naval Electronics Laboratory Center, located at San Diego, California, has been retained to assist in the overall project and has undertaken a study of current methods that could be applied in developing the combat simulation model. We have begun Phase I. The research plan, that is the output of this phase, will be developed by an outside contractor. We have high hopes of being well along in this effort by the end of the year.

We have now seen the need for a new system with which to train tactical decision makers and how, by using simulation techniques, such a system, known by the acronym TESE, is being developed. Further, we have examined the progress that has been made in this development thus far.

There are dangers and risks involved in employing simulations for educational purposes. Assumptions made during model construction must be explicitly available to the student. He must know which variables are being manipulated and what assumptions have been made about the interaction of variables to reduce them to mathematical terms in the model. Otherwise, he will not be able to interpret the results he achieves. We don't want false learning, and this is a real danger in this educational approach. But we believe that by exposing the model to a wide range of intuitive judgment in the academic environment we can validate the techniques and assumptions used in its construction to a useful level of confidence.

Although, undoubtedly, there are pitfalls in attempting as complicated a development as the TESE project we have attempted to minimize them with a phased development prosecuted by a combination of educators and technicians. The rewards that can be realized from the TESE are great and we have high hopes of success.

Measuring Communicative Skills

H. William Greenup
Education Center
Marine Corps Development and Education Command
Quantico, Virginia 22134

INTRODUCTION

To be able to speak, to hear, to read, and to write, is to participate in and profit by the greatest of human achievements: the communication of mankind's experiences so that each succeeding generation can go on from where the preceding one left off. No generation in history has been more aware of the importance of communications than this one, especially in the United States. We have devoted tremendous sums of time, money and energy to improving our communicative skills. Many 20th century scholars have concentrated on measuring our knowledge about the various aspects of the communication process. But despite the work of people like S. I. Hayakawa on semantics, Dr. Ralph Nichols on listening, Miles Tinker and Russell Stauffer on reading, Rudolph Flesch and William Strunk on writing, we still do not communicate gracefully or effectively.

On the whole, our writing is dull, passive and full of meaningless, portentous words and psuedo-scientific jargon; our speech is imprecise and uninspiring; we hear, but we do not listen; we do not read enough of the right things, and when we do read, we read slowly and inefficiently. Rightly or wrongly, many experts ascribe the majority of our social ills to poor communications. Most of us agree wholeheartedly, but plunge right on, oblivious to our own shortcomings and indifferent to the mistakes of others. The situation reminds me of a scene from the film, "Cool Hand Luke," in which the callous warden of a prision road camp, having watched impassively as a prisoner was brutally beaten by a guard for a minor infraction, turns to the other inmates and flatly intones, "What we have here is a failure to communicate." The warden obviously did not understand the meaning of communication. As educators and trainers, we must not similarly misunderstand the function of communication in our society. If, indeed, what we have in the military services is a failure to communicate, we must face the problem squarely and openly.

BACKGROUND

The first thing we must do is to learn more about the communicative process itself: what skills are involved; how these skills relate to one another; how we can devise better methods for developing these skills in the students who pass through our institutions; and, most importantly, how we can motivate these students to continue the process on their own after they have left the classroom.

A wide range of skills, both psycho-motor and cognitive, are involved in the communicative process. For the purposes of this study, however, we will consider them in their broadest sense, the four principal communicative skills: listening, speaking, reading and writing. It is generally acknowledged by educators and psychologists that the four functions are interrelated. Nevertheless, they are taught as separate subjects, and relatively little research has gone into how the development of one ability affects the other three. This paper describes an attempt to take a closer look at the interrelationship between reading and writing.

METHOD

PURPOSE

The Marine Corps Command and Staff College at Quantico, Virginia is a high level school whose mission is to prepare officers of the ranks of major and lieutenant colonel for command and staff duty appropriate to the grade of colonel. Included in its curriculum are over 100 hours of effective communication courses. Since the military officer is expected to be proficient in all of the communicative skills, up to last year every student was assigned an individual research project (IRP) as part of the course in writing. Many, if not most, of the research papers were badly written and contributed little to the program. In 1970, the class was given the option of taking a course in oral and written communication or writing a formal research paper (IRP). Students, some of them still not skilled in writing, elected to work on the research paper because they felt they had something to say. Others, including some naturally gifted writers, chose to take the writing course where they received tutorial guidance and were assigned challenging, but less ambitious writing projects.

During the balance of the 1970-71 academic year, as the faculty evaluated the revised writing course, it was

observed that many of the students with writing problems were also slow readers, or lacked the appetite for reading usually evidenced by the better students. Since certain writers in the field of reading improvement, such as Tinker, and some writing teachers, such as Dr. Joseph M. Woods of Northeastern University, see a definite correlation between reading effectiveness and writing ability, we decided to measure the reading skills as well as the writing skills of the 1971-72 Command and Staff College Class.

PROCEDURE

The study group was composed of 118 officers with varied educational backgrounds: 23 had advanced degrees, 81 had bachelor's degrees, and 14 did not have any college degree. Interestingly enough, subsequent test results did not show a significant positive correlation between the possession of a degree and the ability to read effectively or write clearly.

During the first week of the term the study group was given two tests: the Military Officer Records Examination (MORE) and a 100-item grammar and punctuation test. The MORE is actually the aptitude portion of a discontinued edition of the Graduate Records Examination. The Marine Corps Education Center uses the test under a contractual agreement with the Educational Testing Service. The MORE is a three-hour test of general scholastic ability at the graduate level. It measures the basic verbal and mathematical abilities that a person has acquired over many years. Since mathematical ability was not germane to our study, the group took only the first two sections of the test. Section One includes 60 verbal reasoning questions which provide a good measurement for vocabulary, the key element of verbal ability. Section Two had 40 reading comprehension questions concerning excerpts from various types of prose compositions. We considered the MORE to be a good device for our purposes even though it is not included in lists of standardized reading tests. We had used the MORE previously and believed it to be both a reliable and valid predictor of verbal skills essential to both reading and writing. The MORE has an excellent reliability coefficient of .93. Its content validity lies in its testing indirectly the kinds of skills and abilities that are part of the learning requirements for students in the Command and Staff College -- the ability to read with comprehension, think logically, and see relationships between words and ideas. Although the MORE provided two very useful measures of reading skill and one of writing ability, it alone was not sufficient to guage the students' overall capacity to read effectively and write clearly.

Dr. Argus Tresidder, Professor of English at the Command and Staff College, saw a need for a test to determine how much the students had forgotten about the mechanics of writing -- spelling, punctuation, sentence construction, etc. -- since their last English grammar course. He designed a 100-item test to identify the students' understanding of the basic principles of good writing. These two tests gave us an appraisal of each student's verbal knowledge, plus his ability to apply that knowledge in representative reading and writing tasks. However, we still were not able to determine if a student was an efficient reader.

An efficient reader is one who is able to adjust his reading speed to the best rate for achieving the purpose of the reading and for the kind of material read. For example, someone reading a novel should read much faster with less concern for fact retention than someone reading a psychology text. The mark of the efficient reader is flexibility. The usual measure of reading capability is words-per-minute (WPM). Dr. Russell G. Stauffer, Director of the Reading Study Center at the University of Delaware, has pointed out that WPM is merely a measure of the speed with which an individual recognizes words. He maintains that it is almost completely meaningless in measuring flexibility of reading, or mature efficient reading. Stauffer, therefore, devised the Reading Efficiency Index (REI) which represents the product of reading rate times comprehension accuracy. The determination of the REI can be shown in the following example. Suppose a person reads a passage of 1,000 words in four minutes, and then is able to recall 70% of the key facts contained in the passage. If we multiply his WPM of 250 (which, incidentally, is approximately the rate at which the average reader reads) by his rate of accuracy, .70, we get an REI of 175. According to Stauffer, 175 borders on poor reading (the REI scale is shown in figure 4). Since the REI seemed to be a useful measure of efficient reading, we had the group take a pretest designed by Stauffer.

Finally, we attempted to appraise the students' reading interests, tastes, and attitudes. It is not easy to measure such subjective qualities, but, as Tinker has pointed out, if the teacher is to exercise guidance in reading improvement, appraisal of these factors is essential. We devised a 21-item questionnaire that sought to determine such things as:

1. The students' general attitude toward reading;

2. How much and what kinds of material they read;

3. What types of reading they prefer;

4. To what extent reading acts as an influence in their lives; and

5. How they rate themselves as efficient readers.

The data collected from these four sources were analyzed to determine if a relationship existed between an individual's reading habits and his reading comprehension. We also propose to explore the relationship between the student's reading comprehension and his ability to express himself clearly and effectively in writing.

RESULTS AND ANALYSIS

Scores on the MORE are reported on a scale ranging from 200 to 900. Each student's performance is summarized in a three-digit number, or score. This number by itself has no interpretable meaning. It derives meaning only from the score scale, which relates it to the performance of other people who have taken the test. Using scoring scales provided by the Educational Testing Service, we were able to compare the performance of individual students with the overall performance of their classmates, and with the national norms for the total population taking the test over the past three years. The overall results of the MORE are shown in Figure 1.

Figure 1

MORE Results for CSC 1971-72

Individual High Score	720
Class Median	500
Class Mean	486 (National Mean is 512)
Individual Low Score	300

The scores were generally favorable. Although 58.5% (69) of the 118 who took the test, scored below the national mean, the class mean compared very favorably with scores available from other high-level service schools, such as the Army Command and General Staff College and the Naval War College's School of Naval Command and Staff. The national mean of 512 is the average score made on the verbal aptitude portion of the Graduate Record Examination (GRE). Since the average officer in our group was 37.8 years old and had not been involved in a formal educational experience for the past 6 years, we really did not expect him to score as high as the generally younger, more academically oriented individual who takes the GRE.

The results of the Grammar and Punctuation Test (G-P) are shown in Figure 2.

Figure 2

G-P Results for CSC 1971-72

Individual High Score	92
Class Mean	50
Class Median	49
Individual Low Score	27

Since we had no national scale with which to compare results, Dr. TRESIDDER drew on his many years of teaching experience and set 55% as the "passing" grade. The group's performance on the G-P was not nearly as encouraging as it had been on the MORE. Only 33 of the 118 students scored 55% or better. Of the 85 who "failed," 22 had grades of 40 or below. The results of the MORE and the G-P indicated that although the class as a whole had the basic verbal capacity to be effective communicators, they were in serious need of work in elementary English composition and practice in writing.

The results of the Reading Efficiency Test, shown in Figure 3, showed the class to have a Reading Efficiency Index (REI) of 275 which put them in the "Above Average" category on Stauffer's Reading Performance Index (see Figure 4).

Figure 3

Reading Efficiency Test Results for CSC 1971-72

	REI
Individual High Score	525
Class Mean	275*
Class Median	259
Individual Low Score	130

*Computed on the basis of average WPM of 344 and average comprehension of 80%.

Figure 4

General Reading Performance Index

Rating	REI
Excellent	601 and Above
Good	301 - 600
Above Average	201 - 300
Average	175 - 200
Poor	Below 175

Although the results of the MORE, the G-P, and the REI tests were individually revealing and useful, we were more interested in how the students' aptitudes for vocabulary, verbal reasoning, and reading comprehension compared with their understanding of the basic rules of good writing. We wanted to determine if there was a positive correlation or if these were essentially independent skills.

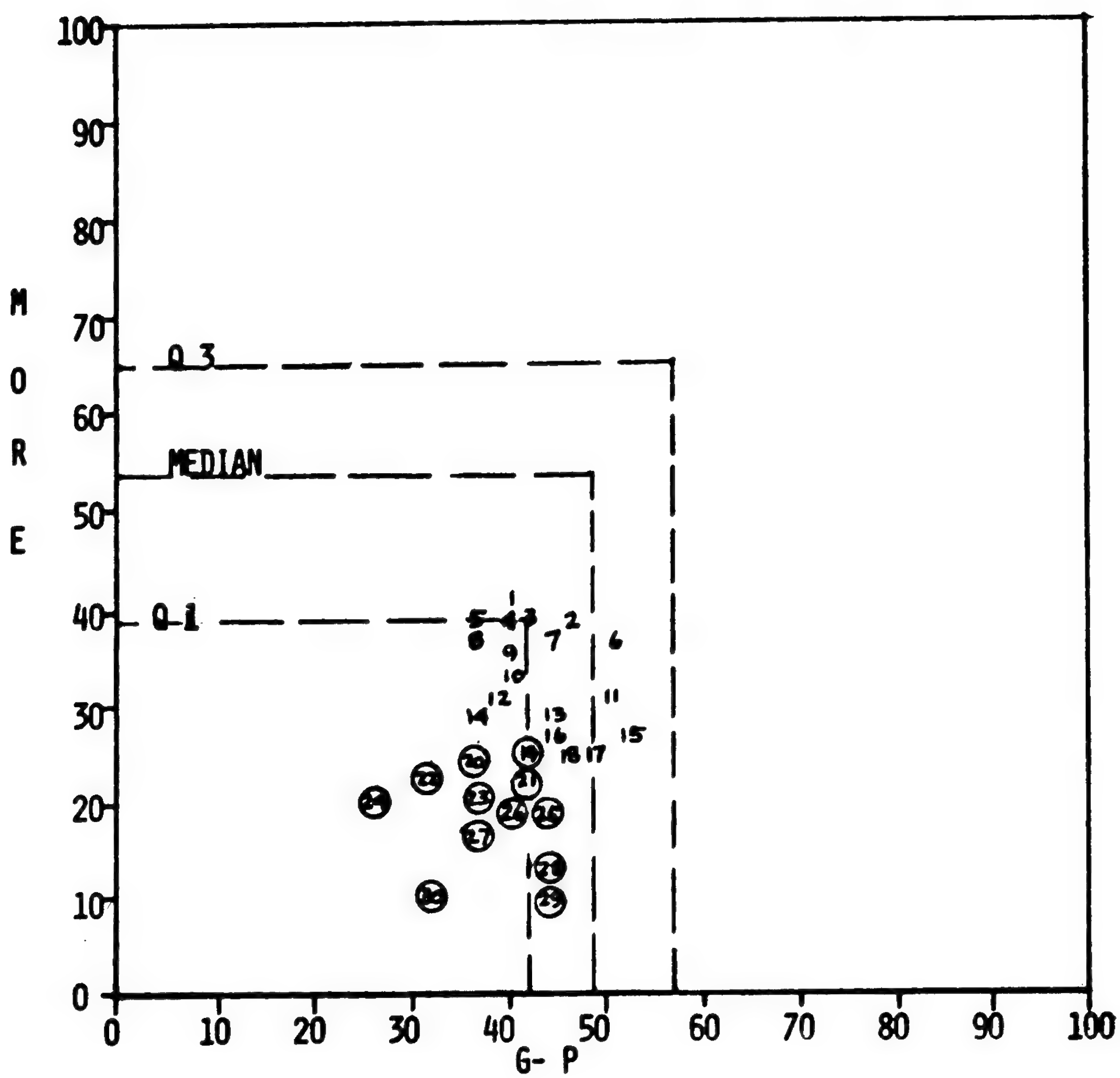
Since the MORE measures skills which are essential in both reading and writing, we identified the 30 people who scored in the lower 25% of the class on the MORE, and compared their performance on the MORE with their performances on both the G-P and the REI tests. Since at this stage of our study we needed only an estimate of the correlation between this group's performance on the three tests, we plotted the variables on scattergrams which are shown in Figures 5 and 6.

Figure 5 shows the relationship between the scores of the 30 people who received the lowest scores on the MORE and their scores on the G-P. A strong positive correlation is evident. Eighteen of the 30, or 60%, scored within the 25th quartile. An even stronger correlation is evident in the cases of the botton 10% of the class, which are circled. Ten out of 12, or 83%, fell within the 25th quartile.

Figure 6 shows the relationship between the scores on the MORE and the REI test of the same 30 people. Here the correlation is even more striking. Twenty-three of the group, or 77%, scored in the lower quartile on the Reading Test. Eleven of the 12 who comprised the bottom 10% on the MORE were also at the bottom of the reading test.

FIGURE 5

Relationship Between the Scores of People Scoring Low on the MORE and Their Scores on the G-P



PTEP PERSONNEL DATA

NAME 001 SSAN 002 RATE AND DOR 003 PRATE AND DOR 004 PNEC AND DATE 005 SNEC AND DATE 005

TNEC AND DATE 007 TPC 009 POL PAT 010 POS PAT 011 DUTY STA AND DROB 012 INST EXP MO 013 CONST/CONV MO 014 CIV ED YR 015

PRD 016	EAOS 017	SSQUAL/REQUAL 018	DUTY STA HISTORY 030/034	BASIC BATTERY SCORES 035/047		CORR/OFF DUTY CS AND DOC 050/054
			030	035	AFQT	050
			031	036	GCT	051
			032	037	ARI	052
			033	038	MECH	053
			034	039	CLER	
				040	SPT	
				041	SONAR	
				042	RADIO	
				043	ETST	
				044	READ	
				045	MATH	
				046	PHYS	
				047	ELEC	

COURSE COMP AND DOC 055/064	SSWB COMP 065/084	CAT RECORD 100/119	SAT RECORD 120/129
	SSWB VOL PART DOC	TEST NO DATE SCORE	TEST NO DATE SCORE
055	065	100	120
056	066	101	121
057	067	102	122
058	068	103	123
059	069	104	124
060	070	105	125
061	071	106	126
062	072	107	127
063	073	108	128
064	074	109	129
	075	110	
	076	111	
	077	112	
	078	113	
	079	114	
	080	115	
	081	116	
	082	117	
	083	118	
	084	119	

FIGURE 11

How well a person learns what is taught and how much he will read on his own depend largely upon his interests. As Tinker and Bond have pointed out, interests provide the motivation that induces the individual to respond eagerly to various activities, including reading. The questionnaire on reading habits, used in this study, represented our initial attempt to determine the current breadth and strength of the students' interests and their level of tastes. "Breadth of interest", according to Tinker, "is indicated by the varieties of reading activity taking place. The strength of our interest pattern is indicated by the time and effort devoted to different types of reading material. Standards of taste in reading are highly subjective. But, when the level of taste of an individual is discovered, it is possible to note whether added experience and teacher guidance lead to the reading of 'better' books." It was with these points in mind that we designed our questionnaire.

Completed questionnaires were received from 96 officers -- 81.3% of the study group. Responses indicated that the average officer spends 20 hours reading each week. Of these 20 hours, 13 (65%) are devoted to work-related reading and 7 to pleasure reading. Of the 7 hours spent in reading for pleasure, 70% is devoted to reading non-fiction material. Newspapers and magazines comprised the bulk of the non-fiction material read, while books were divided about evenly between fiction and non-fiction. The average student estimated that he had read 11 fictional and 10 non-fictional books in the past year, resulting in a surprisingly high total of 21.

In the field of fiction, the students were asked to rate their preferences among such types of books as current best sellers, mystery/adventure, war stories, historical novels, drama, and poetry. Historical novels and current best sellers proved to be the most popular types. Historical novels were rated as their first or second choice by 58% of the group, while 56% marked current best sellers in the top two categories. Drama and poetry, as might be expected, were the least popular subjects.

The students again showed a definite interest in history when asked to rate their preferences for non-fiction among such topics as psychology, political science, sociology, economics, history, military strategy, and science. In this category, history was rated as their first or second choice by 66% of the class. The next most popular topic was military strategy, which ranked first or second among 44% of the students. One interesting point, in light of the military services' search for answers to the problems of race relations, drug abuse, and discipline, was the low preference for

psychology and sociology. Only 7% expressed a significant interest in the subject of psychology, while 13% rated sociology as a preferred topic. There is nothing significant in these figures themselves, but they do indicate an area that may warrant more attention.

The officers were asked which media they used to keep abreast of current affairs, and, of those indicated, which one they used most frequently. The results are shown in Figure 7.

Figure 7

Use of Media for Current Affairs Information by CSC 1971-72

Media	Percentage Who Use	Percentage Who Use Most Frequently
Newspapers	94.8	57.3
News Magazines	86.5	22.9
Commentary Magazines (Harpers, Sat. Review, etc.)	20.8	1.0
Books	33.3	--
Radio	66.7	4.2
Television	91.7	12.5
Conversation	45.8	2.1

The students were also asked to indicate how frequently they read eleven different types of publications. The percentages of officers who indicated they read these publications regularly were as follows:

Type of Publication	Percentage Who Read Regularly
One daily newspaper (Two or more daily newspapers - 19%)	95
One weekly newsmagazine (Two or more weekly newsmagazines - 29%)	63
Commentary newsmagazines (Harper, Sat. Review, etc.)	8

Type of Publication	Percentage Who Read Regularly
Special Interest Magazines (Outdoor Life, National Geographic, etc.)	45
Business Magazines	18
Technical or Professional Magazines: Nonmilitary	18
Technical or Professional Magazines: Military	80
Military News Publications (Armed Forces Journal, Navy Times, etc.)	58
Entertainment Magazines (Playboy, Sports Illus., etc.)	35

In order to accurately assess a person's attitudes and understand why he is or is not motivated toward a particular activity, it is necessary to learn something about his self-image in regard to that activity. With this in mind, we asked the students to rate themselves as readers. The results are shown in Figure 8.

Figure 8

Self-Assessment of Reading Skills by CSC 1971-72
(Expressed in percentages of students responding)

	Speed	Retention
Very Good	12.5	8.3
Good	38.5	46.9
Fair	39.6	40.6
Poor	9.4	4.2

Our analysis of the questionnaires indicates that the average student in the Command and Staff College has a positive attitude toward reading and his abilities as a reader. He likes to read. His principal fields of interest are history and professional military subjects, but he spends an appreciable amount of time reading on a wide range of general information topics. He devotes most of his reading time to work-related and factual material, yet still manages to read a fair amount of modern fiction. He uses more than one type of media to obtain information on current affairs, but relies most heavily on newspapers and magazines.

SUMMARY

The military officer is expected to be proficient in the four principal communicative skills: listening, speaking, reading, and writing. It is generally acknowledged that the four functions are interrelated. Nevertheless, they are taught as separate subjects, and relatively little research has gone into how the development of one ability affects the other three. In this study, we are seeking to determine if a significant relationship exists between an individual's reading and writing ability. We began by administering the Military Officer Records Examination (MORE), a Grammar-Punctuation Test, and a Reading Efficiency Test to 118 field grade officers attending the Marine Corps Command and Staff College. The results of these three tests indicate a strong correlation between an individual's verbal skills, his reading comprehension, and his understanding of the basic principles of good writing. Using a survey questionnaire, we also explored the relationship between the student's reading habits and his performance on the three tests. We found that the average student at the Command and Staff College has a positive attitude toward reading, his interests covered a broad range of topics, and his level of tastes were generally high. We found evidence that a student who claimed he did not like to read usually lacked confidence in his ability as a reader. As a rule, the same student read less, and his interests were more limited.

This study is still in progress. We need to observe many more examples of the students' writing before any conclusion can be reached about the strength of the relationship between reading and writing. However, our efforts to date indicate a relationship does exist, and that the key to success in both reading and writing is practice, practice, and more practice.

REFERENCES

1. Bond, G. L. and Tinker, M. A. Reading Difficulties, Appleton-Century-Crofts, 1957.

2. Cole, Tom J. "College Teaching of Reading: The Literature", Improving College and University Teaching, Winter, 1971. 46.

3. Guide to the Use of GRE Scores in Graduate Admissions 1970-1, Educational Testing Service, Princeton, N. J., 1970.

4. Gagnè, Robert M. The Conditions of Learning (2nd Ed.), Holt, Rinehart and Winston, Inc., New York, 1970.

5. Hayakawa, S. I. Language in Thought and Action, Harcourt, Brace, and Co., New York, 1949.

6. Kavanagh, James F. (Ed.) Communicating by Language: The Reading Process, National Institutes of Health, U.S. Department of Health, Education, and Welfare, Bethesda, Md., 1968.

7. Liebert, Robert E. (Ed.) Diagnostic Viewpoints in Reading, International Reading Association, Inc., Newark, Del., 1971.

8. Palmer, William S. "Reading, Writing, and the Realm of Reason," Phi Delta Kappan, April, 1971. 473.

9. Stauffer, Russell G. Directing Reading Maturity as a Cognitive Process, Harper and Row, New York, 1968.

10. Tinker, Miles A. Bases for Effective Reading, University of Minnesota Press, Minneapolis, 1965.

READING HABITS SURVEY

Throughout your career as a Marine officer, you have frequently been called upon to demonstrate proficiency in each of the four principal communicative skills: listening, speaking, reading, and writing. Educators generally acknowledge that these four functions are interrelated. Nevertheless, they are taught as separate subjects, and relatively little research has gone into how the development of one ability affects the other three.

As part of a continuing effort to assist the individual student and to improve the teaching of communications throughout the Education Center, the Academic Department of the Education Center is conducting research into the interrelationships between reading and writing. The study first seeks to determine if a significant relationship exists between an individual's reading habits and his degree of reading comprehension. It will also explore the relationship between the individual's reading comprehension and his ability to write clearly and effectively.

The attached questionnaire is designed to obtain information about the reading habits of the students in the Command and Staff College. This information will be used to construct a group profile of the class's reading habits. The reading profile will later be compared with a corresponding writing profile to determine if a relationship does in fact exist between the two. The study is <u>not</u> a Command and Staff College project. It is intended to collect raw data for research purposes only.

Please return completed questionnaires to the student drop box outside Room 104 by 1300 on Friday, 10 September.

READING PROFILE

1. NAME______________________________

2. RANK______________________________

3. MOS_______________________________

4. Number of years since last formal school experience. (This includes civilian or military schools lasting longer than two weeks.) ______________________________

5. What was your last full-time assignment before you reported to Command and Staff College? ____________________
____

6. As a rule, do you enjoy reading?____________________

7. Approximately how many hours a week do you spend in work-related reading? ____________________

8. Approximately how many hours a week do you spend reading for pleasure? ____________________

9. What percentage of this time is spend in reading:

fiction ____________

non-fiction ________

10. Do you have a personal library? ________ If so, approximately how many books are in your collection? ____________

11. Do you read book condensations in preference to full-length books?___________

12. Approximately how many books have you read in the past year?

fiction ____________

non-fiction__________

13. What was the name and author of the last book you read?

____

14. What dictionary do you own? ___________________________

15. In the field of fiction, rate your preferences from one to seven.

a. Current best sellers __________

b. Mystery/adventure __________

c. War stories __________

d. Historical novels __________

e. Drama __________

f. Poetry __________

g. Other ______________________________
(Please List)

16. In the field of non-fiction, rate your preferences from one to eight.

a. Psychology __________

b. Political Science __________

c. Sociology __________

d. Economics __________

e. History __________

16. (Continued)

f. Military Strategy ______

g. Science ______

h. Other__________ ______
(Please List)

17. What media do you rely upon to keep abreast of current affairs? (Mark an X in the appropriate blanks. Then, please circle the medium you use most frequently.)

a. Newspapers ______

b. News Magazines ______

c. Commentary Magazines (Harpers, Sat. Review) ______

d. Books ______

e. Radio ______

f. Television ______

g. Conversation ______

18. How often do you read the following types of publications? (Place an X in the appropriate space. Then, please circle those to which you personally subscribe.)

	REGULARLY	OCCASIONALLY	RARELY	NEVER
a. One daily newspaper	______	______	______	______
b. Two daily newspapers	______	______	______	______
c. One weekly newsmagazine	______	______	______	______

18. (Continued)

	REGULARLY	OCCASIONALLY	RARELY	NEVER
d. Two or more weekly news-magazines	____	____	____	____
e. Commentary Magazines (Harpers, Saturday Review, etc.)	____	____	____	____
f. Special interest magazines (Outdoor Life, National Geographic, etc.)	____	____	____	____
g. Business Magazines (Fortune, Forbes, etc.)	____	____	____	____
h. Technical or professional magazines: Nonmilitary (Foreign Affairs, Educational Technology, etc.)	____	____	____	____
i. Technical or professional magazines: Military (Marine Corps Gazette, Naval War College Review, etc.)	____	____	____	____

18. (Continued)

	REGULARLY	OCCASIONALLY	RARELY	NEVER
j. Military news publications (Armed Forces Journal, Navy Times, etc.)	______	______	______	______
k. Entertainment Magazines (Playboy, Sports Illus., etc.)	______	______	______	______

19. Have you ever taken a reading course?______ If so, when? (give year) __________ What was the name of the course?______________________________
Do you believe it improved your reading ability?________

20. How would you rate yourself as a reader?

______ Too fast ______ Fast ______ About right ______ Slow ______ Too Slow

21. How would you rate your ability to retain what you read?

______ Very good ______ Good ______ Fair ______ Poor

AUTOMATIC INTERACTION DETECTION AMONG VARIABLES IN PERSONNEL EVALUATION

By

Janos B. Koplyay
Personnel Research Division
Air Force Human Resources Laboratory (AFSC)
Lackland Air Force Base, Texas

Multiple regression analysis is a powerful approach to the formulation and the analysis of research problems, and the testing of hypotheses. It is less restrictive than multiple correlational analysis; e.g., multiple regression analysis does not assume that the predictor variables constitute a multivariate normal distribution. The absence of this restriction permits the introduction of categorical predictor variables. One use for such variables is the establishment of mutually exclusive groups and the testing of the hypothesis that knowledge of group membership at different levels of a predictor variable improves the accuracy of prediction of a criterion of interest. The automatic interaction detector improves the power and efficiency of the application of multiple regression analysis through the identification of optimal configurations of predictor variables for criterion prediction. Joint familiarity with regression techniques and the application of the automatic interaction detector will provide the research scientist with an effective tool. Without the automatic interaction detector, the establishment of optimally effective sets of predictor variables is essentially a cut-and-try, guesswork process. With automatic interaction detection, guidance is offered directly as to the optimal prediction possible with the predictor set, and the identification of

reduced subsets of predictors which most closely approximate the total validity of the full set of predictors. In this sense, AID-4 is a model identifying process.

The multiple regression technique as illustrated by Bottenberg and Ward (1963), starts with a K-category full regression model including all the predictor variables (categorical and/or continuous) and the basic procedure consists of testing for the significance of the difference between the error sum of squares resulting when some of the least-square weighted categorical memberships are not taken into account in the (K-n)-category restricted model where n is the number of restrictions imposed upon the full model. The test of significance is done by the F-statistic, comparing the minimized error sum of squares of the full model with that of the restricted model. This comparison indicates the extent to which the eliminated n categorical memberships contributed to the accuracy of predicting the criterion variable.

For a simple example, let us suppose that we have two predictor variables $x^{(1)}$ with three levels, i.e., high school degree, undergraduate degree and graduate degree; and $x^{(2)}$ with two levels, i.e., pilot or navigator. The criterion variable is some test score on a 50-item test and we have 60 individuals in the experiment. (The actual data was taken from an example in Hays' Statistics, Holt, Rinehart and Winston, 1963, p. 403.) The simple two predictor, one criterion multiple regression model is:

$$\underline{\text{Model 1}} \qquad y_1 = a_0 u + a_1 x^{(1)} + a_2 x^{(2)} + e_1$$

which after the conventional multiple regression yields a solution of R_1^2 = .7508 and a mimimized error sum of squares of q_1 = 1607.4670.

Testing for interaction one would include a product term in the model:

$$\underline{\text{Model 2}} \qquad y_2 = b_0 u + b_1 x^{(1)} + b_2 x^{(2)} + b_3 x^{(1)} \cdot x^{(2)} + e_2$$

Model 2 is the so called "full model" and Model 1 is the "restricted model." It is restricted because we impose the restriction of $b_3 = 0$ upon Model 2 thus obtaining Model 1. By comparing the minimized error sums of squares of Model 1 and Model 2, q_1 and q_2 respectively, one gets an indication of the contribution of the product term (or "interaction") to the predictive efficiency of the system. The solution of Model 2 gives an R_2^2 = .8184 and q_2 = 1171.8683.

The F-statistic is computed by:

$$F = \frac{(q_1 - q_2)/(4 - 3)}{q_2/(60 - 4)} = 20.82$$

with df = 1 and 56. We can make further "guesses" about the predictor variables. Let us assume that predictor $x^{(1)}$ has a quadratic component and that the previously hypothesized interaction is also present. Our model will look like:

$$\underline{\text{Model 3}} \qquad y_3 = c_0 u + c_1 x^{(1)} + c_2 x^{(2)} + c_3 x^{(1)} \cdot x^{(2)} + c_4 \cdot \left[x^{(1)}\right]^2 + e_3$$

The solution of Model 3 yields an R_3^2 = .8423 and a minimized error sum of squares q_3 = 1017.7627. The F-statistic is:

$$F = \frac{(q_2 - q_3)/(5 - 4)}{q_3/(60 - 5)} = 8.33$$

with df = 1 and 55. Additional possible models are listed below:

<u>Model 4</u> $y_4 = d_0u + d_1x^{(1)} + d_2x^{(2)} + d_3x^{(1)}x^{(2)} + d_4\left[x^{(2)}\right]^2 + e_4$

$R_4^2 = .8184 \qquad q_4 = 1171.8683$

<u>Model 5</u> $y_5 = k_0u + k_1x^{(1)} + k_2x^{(2)} + k_3x^{(1)}x^{(2)} + k_4\left[x^{(1)}\right]^2 + k_5\left[x^{(2)}\right]^2 + e_5$

$R_5^2 = .8423 \qquad q_5 = 1017.7627$

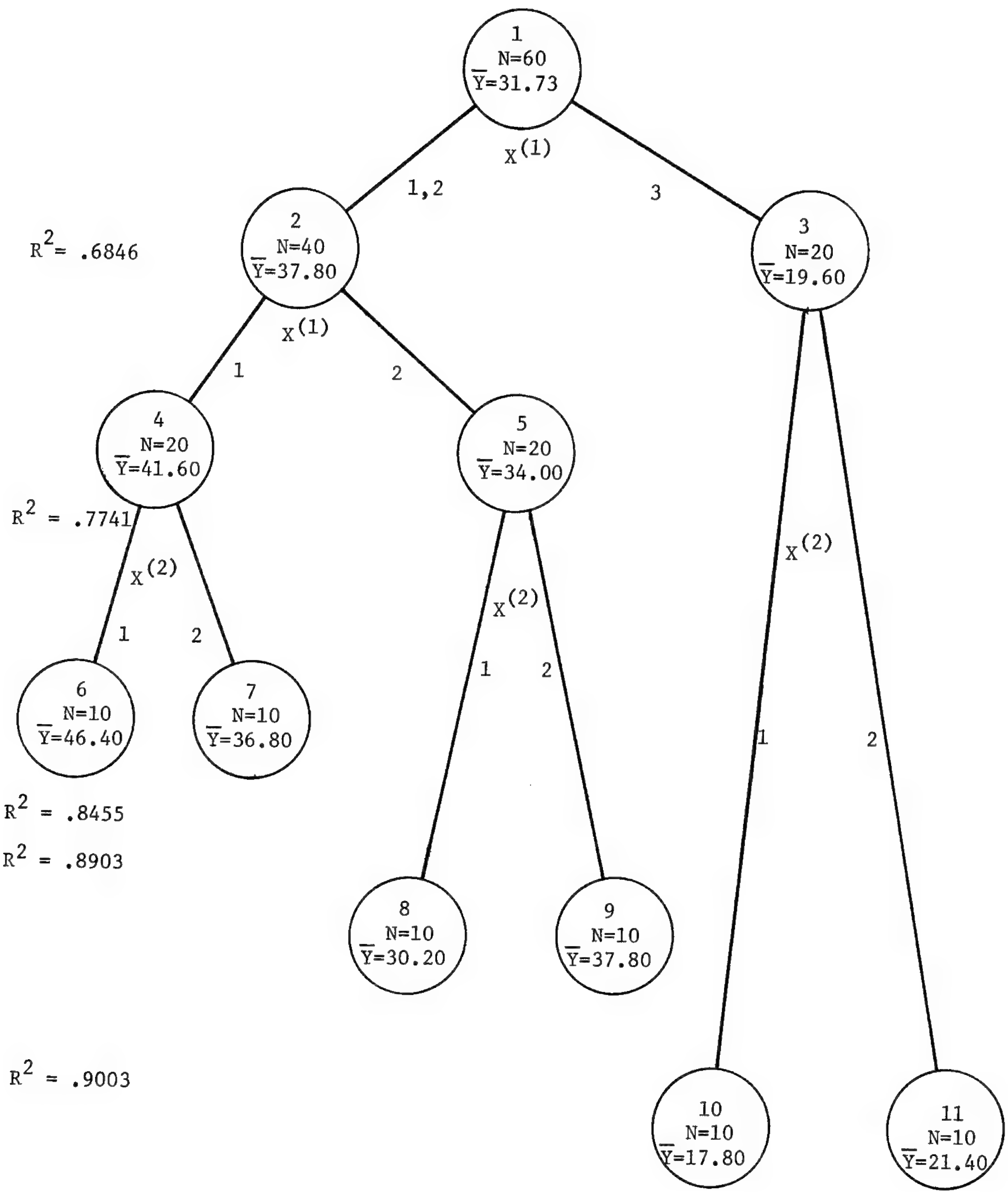
It should be obvious at this point that had we had a more complex problem, for example 40 predictor variables with 10 levels each, the guesswork would have been futile and totally unreasonable. The number of possible mutually exclusive categories in the model would be 10^{40}, most of which would be empty, considering that the total population of the earth is approximately 4×10^9.

This was the reason for implementing and developing AFHRL's version of AID-4. The algorithm of AID-4 is a reversal of the model building process. Rather than starting with a full model, including all possible predictors and their simple and complex interactions, AID-4 starts with the ultimate restricted model, namely, the whole group as a unit. By a unique splitting process maximizing the between sum of squares (BSS) for the categories of each variable while minimizing the error sum of squares (within group sum of squares) AID-4 seeks out that variable which has the largest BSS and splits the original group into two mutually exclusive groups on this variable at that category where the maximum BSS occurred. For example, given an 80 variable problem with 10 categories per variable, if the maximum BSS was found in Variable 9 and between categories 1, 2, 3 and 4, 5, 6, 7, 8, 9, 10; the original Group 1 will be split into two

mutually exclusive groups: (a) Group 2 consisting of those individuals whose response to Variable 9 was 1 or 2 or 3, and (b) Group 3 consisting of the remainder of the individuals whose response to Variable 9 was 4, 5, 6, 7, 8, 9 or 10. In actuality, AID-4 has identified the first level full model consisting of 2 groups. The test of significance is an F-test comparing the minimized error sum of squares of the full model (2 groups) and the restricted model (original 1 group). The test of significance for the first split is equivalent to an F-test obtained by a one-way analysis of variance comparing the 2 groups on the criterion variable. The process continues until a specified stop-criterion is reached. Each time a split occurs, the resulting j mutually exclusive groups represent the full model, and the minimized error sum of squares of this model is compared with the error sum of squares of the previous model, consisting of (j-1) mutually exclusive groups. The final split represents an optimal full model which could have been hypothesized before starting to impose restrictions. Going from the final model with the last split towards the original unsplit group, each unsplit group represents an additional restriction.

For our example, the AID-4 splitting process is illustrated in Figure 1. Going down the branches of the tree-pattern, one can identify the simple and complex interactions of the optimum polynomial multiple regression equation. We know that we have predictor variables $x^{(1)}$ and $x^{(2)}$. The first two splits occurred on $x^{(1)}$, $x^{(1)}$ respectively, hence we have an $\left[x^{(1)}\right]^2$ term. The first three splits occurred on $x^{(1)}$, $x^{(1)}$

FIGURE 1. SPLIT DIAGRAM



R^2 = .6846

R^2 = .7741

R^2 = .8455

R^2 = .8903

R^2 = .9003

$x^{(2)}$ respectively, hence we have an $\left[x^{(1)}\right]^2 \cdot x^{(2)}$ term. The second branch from the left is identical to the first identifying the same $\left[x^{(1)}\right]^2 \cdot x^{(2)}$ term. The third branch from the left split on $x^{(1)}$, $x^{(2)}$ respectively, hence we have an $\left[x^{(1)} \cdot x^{(2)}\right]$ term.

Thus, the optimal model is:

Model 6

$$y_6 = p_0 u + p_1 x^{(1)} + p_2 x^{(2)} + p_3 x^{(1)} x^{(2)} + p_4 \left[x^{(1)}\right]^2 + p_5 \left[x^{(1)}\right]^2 \cdot x^{(2)} + e_6$$

which yields, after conventional solution, an $R^2 = .9003$ which is the same as AID-4 arrived at after the final split. Note that Model 6 does not contain a term $\left[x^{(2)}\right]^2$ which is consistent with the previous findings namely that Model 3 and Model 5 were identical (the only difference being that Model 5 contained $\left[x^{(2)}\right]^2$).

The major advantage accruing to the task scientist using AID-4 is obtaining the maximum squared composite correlation without the task of attempting to identify the various relevant combinations of linear and non-linear interaction terms by trial and error necessary in the full model of the multiple regression technique. AID-4 automatically identifies these terms. The means of the final categorical groups are the proper weights to be assigned for each of those groups in predicting the criterion variable. An additional major advantage is that out of a regression analysis with a large number of predictor variables, there may be only a small subset of predictor variables which are of significance in the prediction system. AID-4 identifies such a subset of predictors automatically. Finally, the branching pattern facilitates interpretation of the results. In our sample example, it is much more meaningful to

identify Group 6 on Figure 1 as pilots who have advanced academic degrees and who have a predicted score of 46.40, than in a polynomial regression equation where one would have to square "educational level" and multiply it by "pilotness" in order to identify the term $\left[x^{(1)}\right]^2 \cdot x^{(2)}$. In a large prediction system, attempts to identify and include all possible combinations of interaction terms represents a practical impossibility without the help of AID-4.

Many additional and useful bits of information are provided by the output of AID-4, some of which are: (1) at each split, the increased present total explained variance (R^2) is printed, together with a statistical test of significance for the difference between the error sum of squares of the new model and the previous model prior to the split; (2) the splits occur in a descending order of importance, that is, the first split identifies that variable which contributes the most to the explained variance; the second split identifies the second variable or a subset of the first split as the next most important contributor to the explained variance; and so on. This hierarchy is very helpful especially if after a few splits a reasonably high R^2 is obtained, thus giving the researcher an option of using only a few predictors in the prediction system; (3) the branching pattern of splits reflects trends of characteristics specific to the groups split; that is, it can serve as an "eyeball" pattern analysis. Following the path of each branch of the split-tree, one can identify major characteristics of the final groups on which they differ the most in light of the criterion measure;

(4) cross-validation and double cross-validation options which either splits the original sample into two random samples or takes two given samples, treats each sample separately, determining an optimal split pattern for each and the associated R^2. Then it forces the split pattern of Sample 1 upon Sample 2 and vice-versa computing a squared composite correlation for these forced splits. The differences between the optimal R^2 for each sample and the corresponding squared composite correlation obtained by forced splitting is a good indicator of the stability of the system; (5) selective or "partial" effects of the predictors are identified such that even if the so-called "main effect" of a particular variable in a complex analysis of variance results in a non-significant F-ratio, AID-4 selectively indicates the level on the other variable(s) at which this non-significant effect becomes significant.

Copies of the write-up and program (to be loaded on a tape provided by the user) can be obtained by written request from Dr. Janos Koplyay, Chief, Statistical and Computer Technology Section, AFHRL/PHSM, Lackland AFB, Texas 78236.

References

Bottenberg, R. A., and Ward, J. H. Applied multiple linear regression. PRL-TDR-63-6. Lackland AFB, Tex.: 6570th Personnel Research Laboratory, Aerospace Medical Division, March 1963.

ENLISTED JOB SATISFACTION IN THE AIR FORCE: A Study at the Task Level

By
R. Bruce Gould
Raymond E. Christal
Occupational and Career Development Branch
Personnel Research Division (AFHRL)
Lackland Air Force Base, Texas

The title of this paper was selected not because the authors have succeeded in capturing that elusive construct, job satisfaction, but to acknowledge that a long-range research program has now begun. The work unit is entitled "evaluation of the impact of Air Force work tasks assigned on job satisfaction, felt utilization of talent, and career decisions."

Zero draft, strength and budget reductions, increased training requirements through technological advances, and resulting emphasis on personnel utilization and retention provide the operational requirement for this research; availability of a comprehensive occupational data base provides the means. The research program's purpose is to identify operational implications for selection, classification, assignment, and job structuring actions from data obtained at the performance or task level.

This paper will present three satisfaction indexes which will be used in the research program, and primary attention will be given to the use of one of these scales as a broad indicator of the extent of enlisted job satisfaction in different specialties. A brief examination will be made of four specialties to suggest some of the causes of expressed dissatisfaction. Intensive studies have been undertaken of these specialties but are not yet completed. Before presenting the satisfaction scales, there will be a brief discussion of the Air Force classification structure and the source of the data base.

A five-digit numerical Air Force Specialty Code or AFSC is used in the classification system to differentiate enlisted jobs. The five-digit AFSC is similar in meaning to the MOSs of the U.S. Army and Marine Corps or ratings of the U.S. Navy and Coast Guard. The first three digits and the last digit of the code complete the specific job classification. This classification is referred to as a career ladder. The fourth digit is the skill level of the incumbent and when skill level is not differentiated in comparing career ladders, an "X" is used in place of the fourth digit as will be seen later.

The USAF Job Analysis program has provided the data base. The job analysis procedures were developed and refined more than a decade ago (Morsh, Madden, & Christal, 1961). Similar job survey programs are being conducted by the Canadian Armed Forces and the U. S. Army, Navy, Marine Corps, and Coast Guard. Under the USAF program, inventories are constructed which contain all the tasks conceivably performed in a given career ladder or in several related career ladders in a given specialty. The inventories are sent to 100 percent of the job incumbents in small population career ladders and to proportionally decreasing percentages of stratified random samples in larger career ladders. The job incumbents first complete several information items such as the job satisfaction criteria reported in this paper. They then check each task they perform and rate the performed tasks according to the amount of time spent (Morsh & Archer, 1967).

Since mid 1966, all job inventories have contained job satisfaction scales. During this period, 105 of the 238 Air Force career ladders have been surveyed. This represents the major Air Force enlisted jobs and constitutes a data base of some 100,000 respondents.

YOUR RESPONSES TO THE FOLLOWING THREE ITEMS WILL BE HELD IN STRICT CONFIDENCE AND WILL BE USED FOR RESEARCH PURPOSES ONLY.

I PLAN TO REENLIST:	I FIND MY JOB:	MY JOB UTILIZES MY TALENTS AND TRAINING:
1 ☐ NO, I PLAN TO RETIRE	1 ☐ EXTREMELY DULL	1 ☐ NOT AT ALL
2 ☐ NO, I PLAN TO SEPARATE WITHOUT RETIREMENT BENEFITS	2 ☐ VERY DULL	2 ☐ VERY LITTLE
3 ☐ UNCERTAIN, PROBABLY NO	3 ☐ FAIRLY DULL	3 ☐ FAIRLY WELL
4 ☐ UNCERTAIN, PROBABLY YES	4 ☐ SO-SO	4 ☐ QUITE WELL
5 ☐ YES	5 ☐ FAIRLY INTERESTING	5 ☐ VERY WELL
	6 ☐ VERY INTERESTING	6 ☐ EXCELLENTLY
	7 ☐ EXTREMELY INTERESTING	7 ☐ PERFECTLY

Figure 1 Satisfaction Scales

Figure 1 illustrates the job satisfaction scales as they appear in the Background Information section of current USAF Job Inventories. The scale of primary interest i.e., felt utilization,

appears with two other scales, reenlistment intent and job interest. The reenlistment and job interest scales will not be elaborated on here; however, each appears to provide unique variance to the overall prediction of satisfaction within the Air Force environment. For the purposes of this paper, the operational definition of job satisfaction is the extent to which job incumbents feel that their current jobs utilize their talents and training. The felt utilization scale appears to have the greatest relationship to job performance at the task level and provides a simple, effective means of identifying ladders which need extensive investigation and development of specific operational recommendations. To this extent, the scale may be termed a "troubleshooting" scale.

Respondents to the felt utilization scale indicate the extent to which they feel their job utilizes their talents and training. Response options range from "not at all" to "perfectly" on a 1 to 7 point linear scale. This scale is useful in identifying potential satisfaction problems when respondents are dichotomized into those reporting their utilization as "very little" or "not at all" and those responding "fairly well" to "perfectly." The complete 1 to 7 linear scale is promising for use in correlation and regression studies of individual ladders.

Table 1. **Percent Claiming Talents and Training Are Utilized "very little" or "not at all"**

		N	3 Level	5 Level	7 Level
551X0	Pavements Maintenance	1086	63	54	15
671X3	Disbursement	1354	58	45	N/A
303X3	Auto Tracking Radar Rpm	728	51	39	20
811X0	Security	3617	46	36	9
204X0	Intelligence Opers	826	43	39	20
631X0	Fuel Services	2247	40	34	7
681X0	Data Systems	350	39	39	23
293X0	Radio Operator	1265	38	42	22
291X0	Communication Center	2092	37	31	18
.	.	.			
.	.	.			
.	.	.			
362X4	Telephone Equip. Install/Rpm	838	12	14	12
361X4	Cable Splicer	1055	10	14	7
432X0	Jet Engine Mechanic	1622	9	8	6
305X3	Elec. Computer Rpm	1055	7	14	13
324X0	Precision Meas. Equip. Rpm	1212	6	7	8
342X0	Flight Simulator	743	6	3	2
322X1	Weapons Control Sys. Mech.	1285	5	6	3
908X0	Veterinary	542	2	7	2
982X0	Dental Laboratory	488	0	6	3
612X0	Meatcutter	232	0	5	0

Comparing felt utilization in the 105 career ladders of the data sample, there are large differences, particularly at the skill levels of first-term airmen. Differences are illustrated by comparing the percentages of airmen reporting nonutilization of talents and training. Arranging the ladders on a continuum of lowest to highest felt utilization, the percentage of those feeling unutilized at the semiskilled level ranged from 63 percent in the Pavements Maintenance Ladder to zero percent in the Dental Laboratory and Meatcutter Ladders. Table 1 presents several of the ladders with the highest and lowest felt utilization. The table has column values for each skill level within each ladder. In general, 3-level personnel are E-3s or below with 3 years or less service, 5-level personnel are E-4s or E-5s and may be first- or second-term airmen, and 7-levels are E-6s or E-7s in their third or greater enlistment terms. For all 105 ladders, the average percent feeling unutilized at the 3-, 5-, and 7-levels is 24, 22, and 10 percent respectively. This is interpreted to mean that dissatisfaction with the tasks performed is fairly low for Air Force jobs in general. This does not however disguise the fact that some of the 105 ladders surveyed do exhibit excessive job dissatisfaction.

Within ladder satisfaction differences are apparent. As the skill levels increase, dissatisfaction decreases in most ladders. One explanation is that those who feel unutilized tend to leave the service. Also, as skill levels increase, the tasks performed become more demanding and hence better utilize talents and training.

Specific ladders have high proportions of airmen who feel unutilized. Causative factors differ widely and are essentially unique for each ladder. For personnel feeling most unutilized, factors such as overqualification for the tasks performed or mundane repetitous nature of the tasks themselves appear to account for much of the dissatisfaction. The extent of mundane tasks can be seen from job descriptions of the Pavements Maintenance and Security Police Ladders.

Table 2 is an extract from the task job description of 223, 3-level or semiskilled, Pavements Maintenance personnel. Tasks are arranged in descending order according to reported time spent. The cumulative sum of the average percent time spent by all members, shown in the right hand column, indicates that 25 tasks account for a little over 50 percent of the working time for semiskilled personnel. Twenty-two percent of their working time is spent on gardening tasks such as mowing grass and watering plants; servicing equipment and tools accounts for 13 percent of time spent; and manual labor tasks, such as ditch digging, tearing up pavement, and shoveling snow, account for 14 percent of time spent. The mean number of months in service for this group is 22.2 months--an explanative factor for 63 percent of the members feeling unutilized.

Table 2. **Job Description for 55130 Pavements Maintenance Specialist**

D-Tsk	DUTY/TASK TITLE	Percent of members performing	Average percent time spent by members performing	Cumulative sum of average percent time spent by all members
N 6	Mow or edge grassed areas	74.44	9.50	7.07
L 22	Wash or clean equipment	74.89	5.95	11.52
N 1	Control weed growth	58.74	7.23	15.77
N 8	Trim or remove trees or shrubs	57.85	6.13	19.31
N 3	Plant tree, shrubs, grass, or flowers	50.22	4.60	21.62
N 9	Water or irrigate vegetation	34.53	6.35	23.82
L 19	Service motorized equipment with fuel, oil, coolants, or air	46.19	4.70	25.99
L 6	Clean, lubricate, or sharpen tools	47.98	3.88	27.85
N 2	Fertilize vegetation	39.01	4.49	29.60
L 10	Lubricate operating equipment	39.01	4.46	31.35
H 21	Operate air compressors	55.61	3.06	33.05
L 21	Tighten loose bolts or attachments	43.05	3.73	34.65
I 15	Dig trenches or ditches by hand	43.95	3.23	36.07
F 4	Dump loose construction materials, such as sand or gravel	45.29	3.01	37.44
K 16	Sweep paved surfaces	39.01	3.34	38.74
N 5	Maintain sod beds	19.73	6.49	40.02
F 7	Haul loose construction materials, such as sand or gravel	45.74	2.75	41.28
H 38	Spread gravel or other loose materials	38.57	3.13	42.48
I 44	Use pneumatic equipment to breakup or drill holes in pavement	40.36	2.93	43.67
B 13	Supervise grounds maintenance crews	13.90	8.47	44.85
I 28	Hand tamp paving or pavement base materials	38.57	2.83	45.94
K 13	Remove snow and ice by hand	31.39	3.38	47.00
L 13	Perform equipment operational checks	27.35	3.86	48.06
H 30	Rip or breakup paved surfaces	38.57	2.69	49.09
K 17	Use machinery to remove ice or snow	29.60	3.42	50.11
	. . .	. . .		

Entry-level personnel who attend the Pavements Maintenance Technical School receive training which is very different from the tasks they actually perform on the job. The course curriculum consists of 60-hour blocks of instruction in the areas of: (1) soil and paving material testing and maintaining railways; (2) rigid pavements and prefabricated surfaces and shelters construction and maintenance; and (3) maintaining flexible pavements and performing vegetation control. A 30-hour instruction block is given on the characteristics of soils and chemicals and the handling of explosives. The curriculum is consistent with the classification system description of the Pavements Maintenance Ladder but prepares the trainees for a job they do not perform.

Table 3. **Job Description for 81130 Security Policeman**

D-TSK		DUTY/TASK TITLE	Percent of members performing	Average percent time spent by members performing	Cumulative sum of average percent time spent by all members
E	25	Stand guard mount	83.09	7.22	6.00
E	21	Perform sentry duty	71.92	7.35	11.28
E	19	Perform security area foot patrol	74.21	6.81	16.34
E	6	Control entry into or access within restricted areas	69.48	5.46	20.13
E	5	Challenge or identify unknown persons	75.50	4.96	23.88
E	18	Perform security alert team (Sat) duty	64.18	5.17	27.20
E	7	Defend against real or simulated attacks	63.75	4.81	30.26
E	1	Apprehend or detain intruders	68.34	4.17	33.11
E	8	Escort or guard weapons	56.59	4.80	35.82
M	4	Fire weapons to maintain proficiency	60.03	4.35	38.43
M	2	Clean or lubricate weapon mechanisms or parts	51.43	4.34	40.67
M	3	Field strip weapons	52.15	4.27	42.89
E	20	Perform security area motor patrol	52.58	4.12	45.06
F	19	Operate security police vehicles	44.56	4.43	47.03
M	1	Apply preservatives to weapons	41.26	4.31	48.81
E	12	Issue or inspect visitor restricted area badges or credentials	38.97	4.23	50.46
		.	.		
		.	.		
		.	.		

A similar pattern of a few routine tasks which provide little challenge to the first-term job incumbent is shown in the Security Police Ladder. Forty-six percent of the 3-level incumbents reported that their talents are unutilized. Table 3 is extracted from the combined job description of 698 semiskilled police personnel. While there were 336 tasks in the inventory they completed, 16 tasks accounted for a little over 50 percent of their combined working time. The average months in military service for this group is 16.4. Thirty-one percent of their reported time is spent on duties associated with guard mount or patrolling secure areas and 12 percent of their time handling and maintaining weapons. Inspection of job descriptions of other police personnel indicates only minor job expansion during the entire four years of the first enlistment. Individual job descriptions reveal that for many individuals, more or all of their time is accounted for by even a smaller number of routine tasks.

To demonstrate a specific relationship between the nature of tasks performed and felt utilization, a difference description was generated for 583, 5-level personnel in the 915X0, Medical Materiel Ladder. Twenty-six percent of these airmen reported that their job did not

use their talents and training. For each task, the percent of incumbents performing who felt unutilized was compared to the percent of members who felt well-utilized. The difference between the members in each group performing each task was determined and each task rank ordered according to that difference. Table 4 shows the seven tasks falling at the two extremes of that rank ordering. The difference value in the right hand column is positive if a greater percentage of the members performing the task feel their job utilizes their talents and negative if more members feeling unutilized perform the task. The first task listed is "plan procedures for the requisitioning of materiel." Forty percent of the incumbents who feel well utilized perform this task, while only 14 percent of those feeling unutilized perform the task. In the Medical Materiel Ladder, 5-level personnel who feel unutilized tend to perform warehouse duties, while the more satisfied personnel tend to perform administrative tasks such as planning, editing, and coordinating. From this description, it is evident that there is a relationship of the functional area of the tasks performed to job satisfaction.

Table 4. **91550 Group Difference Description, Medical Material Specialists**

D-TSK	TASK TITLE	Percent performing whose talents well utilized	Percent performing whose talents not well utilized	Percent performing - Difference, Talents well utilized minus not so
A 26	Plan procedures for the requisitioning of materiel	40.07	14.08	25.98
J 14	Review the machine run inventory adjustment document	38.05	15.14	22.91
K 29	Verify unit costs of property items	36.36	13.73	22.63
K 1	Adjust prices of materiel obtained by local purchase	35.69	13.38	22.31
M 15	Review receiving documents	47.47	25.70	21.77
K 5	Edit issue requests	46.80	25.35	21.45
A 7	Coordinate status of issue requests with using activity	46.80	25.70	21.10
	. . .	. . .		
E 38	Segregate incoming shipments for inspection	26.94	40.14	−13.20
E 23	Make deliveries to using activity	44.11	57.39	−13.29
E 24	Mark shipping containers	22.56	35.92	−13.36
E 31	Place location symbols on warehouse bins, racks, or bays	19.19	35.21	−16.02
E 21	Locate and pull stock from storage as directed by delivery slips or other release documents	41.08	60.21	−19.13
E 42	Unload incoming shipments	43.77	63.03	−19.26
E 30	Place items in warehouse bins, racks, or bays	39.39	59.51	−20.11

A different pattern for felt utilization emerges from the accounting and finance ladders. The patterns of mundane task performance effecting felt utilization do not readily emerge in the accounting and

finance ladders. The most apparent relationship appears to be the effect of aptitude and education variables. A recent U.S. Marine Corps study of the same occupational field reported similar results (Van Cleve, 1971).

Table 5. **Analysis of 671XX Ladders**

Percent Feeling Talents are Utilized "Very Little" or "Not at All"

		3-level	5-level
671X1	1967	34%	21%
	1970	40%	36%
671X3	1967	30%	25%
	1970	58%	45%

Aptitude Input, 1966-1970

A-95	2806
A-90	1207
A-85	984
A-80	806
- - - - - - - - - - - - Cutting score	
A-75	147

College Graduates

1966	10%
1967	10%
1968	33%
1969	34%
1970	23%

Expression of Positive Reenlistment Intentions

1970 First Termers in 671XX

High School and below	13.56
College Graduates	3.93

Reenlistment Rates (First Term)

Air Force-Wide	FY69	15.8
	FY70	20.3
671XX	FY69	14.8
	FY70	14.4

Table 5 presents a variety of data summaries for the accounting and finance field. Expressed satisfaction, information on career field input, and reenlistment rates are given. The 671X1, General Accounting, and 671X3, Disbursement Accounting Ladders were surveyed in 1967, and again in 1970. At the skill levels held by first-term airmen, there was a substantial increase in expressed dissatisfaction from 1967 to 1970. The most substantial increase was among the 3-level disbursement personnel where the percentage of personnel feeling unutilized rose from 30 to 58 percent. Much of the rise in job dissatisfaction is seen as related to the career field input during the periods covered by the surveys. From 1966, through 1970, 5,950 airmen entered the accounting field.

The aptitude limits for entry required that personnel score at or above the 80th percentile on the Administrative subtest of the Airman Qualification Examination. More than half the personnel were at the 95th percentile which is the maximum possible score. The educational level of the career field input also increased during this period. In 1966 and 1967, 10 percent of the career field entries had college degrees. In 1969, 34 percent of the entries had college degrees, several with postgraduate credits. As previously indicated, 58 percent of the 3-level procurement personnel felt unutilized in their jobs during the 1970 survey. Of that total 3-level group, 51 percent had bachelor level or higher degrees. Again from the 1970 survey, only 4 percent of the college graduates plan to reenlist, while 14 percent of those with no college background expressed positive reenlistment intentions. Comparing the accounting field to the Air Force wide actual reenlistment rates, in 1969, the ladder's reenlistments were slightly less than the Air Force rate of 15.8 percent. Air Force wide reenlistments increased from 15.8 to 20.3 percent in 1970, while the accounting field decreased from 14.8 to 14.4. In the accounting and finance field we see a high portion of the Air Force's most educated and talented input. The aptitude requirements for the job are set high, but the expressed utilization of talent is among the lowest of the 105 ladders surveyed to date.

In summary, a felt utilization scale can be dichotomized to identify career ladders with potential job dissatisfaction and retention problems from job analysis data. Four career ladders which indicated unacceptable levels of job satisfaction received a cursory evaluation. Certainly routine and mundane tasks or specific undemanding job types within a career ladder contribute to job dissatisfaction. Perhaps such actions as rotating personnel within a ladder and providing more varied job experiences or identifying blocks of unskilled mundane tasks and hiring civilians, much the same as the growing role of the civilian KP, will eliminate some of the problem. Ensuring that formal schools do not overtrain could also be very profitable.

In the case of high skill areas, perhaps some entry aptitude levels are set too high or perhaps minimum and maximum aptitude and education level standards should be set in some fields to limit overqualification problems. However, preliminary regression analyses now being conducted on these ladders indicate that we are not yet ready to make such recommendations.

Using the AID regression techniques presented earlier to this conference by Koplyay (1971), the actual number of variables which affect reported utilization are numerous and their interrelationships complex. Factors such as difficulty and variety of tasks performed; age, aptitude, education, race, and grade of the job incumbent; length of time the incumbent has been on the job, in the ladder, and in the Air Force; present and past jobs performed within the ladder; amount and type of technical training; size and command level of organization; location of base of assignment; number of hours worked per week; number of subordinates supervised; mission of organization or unit; incumbents rated performance level; and perceived competence of immediate supervisors have all been found to contribute significantly to the prediction of felt utilization. Unique contribution of controllable factors is difficult to evaluate because of the complex interactions of this multitude of factors. Certainly no operational recommendations can be made from the superficial results presented earlier until the potential unique contributions of the task and aptitude variables are identified; the AID technique appears very promising toward this end. It is hoped that by next year's conference, the results of implemented recommendations for improving job satisfaction can be reported.

REFERENCES

Koplyay, J. B. The Automatic Interaction Detector (AID-4) as an optimal model building system for regression analysis. Paper presented to the 13th Annual Conference of the Military Testing Association, Washington, D. C., 20-24 September 1971.

Morsh, J. E. & Archer, W. B. Procedural guide for conducting occupational surveys in the United States Air Force. PRL-TR-67-11, AD-664 036. Lackland AFB, Tex.: Personnel Research Laboratory, Aerospace Medical Division, September 1967.

Morsh, J. E., Madden, J. M., & Christal, R. E. Job analysis in the United States Air Force. WADD-TR-61-113, AD-259 389. Lackland AFB, Tex.: Personnel Research Laboratory, Wright Air Development Division, February 1961.

Van Cleve, R. R. Job satisfaction -- a study in utilization of talents and job interest. Paper presented to the Second Annual Psychology in the Air Force Symposium. USAF Academy, Colorado. 20-21 April 1971.

THE SET STUDY

A Research Study of the Self-Evaluation Technique

By

John J. Holden

U. S. Army Ordnance Center and School
Aberdeen Proving Ground, Maryland

THE SET STUDY

A RESEARCH STUDY OF THE

SELF-EVALUATION TECHNIQUE

INTRODUCTION

TODAY I WOULD LIKE TO PRESENT THE FINDINGS OF THE SET STUDY WHICH IS A RESEARCH STUDY OF THE SELF-EVALUATION TECHNIQUE.

PURPOSE OF THE STUDY

THE PURPOSE OF THE SET STUDY WAS TO DETERMINE IF STUDENT SELF-EVALUATION IMPROVED STUDENT PERFORMANCE FOR EIGHT WELDING PROJECTS.

PROCEDURE OF THE STUDY

THE TRAINEES PARTICIPATING IN THE STUDY WERE STUDENT WELDERS ENROLLED IN THE WELDING COURSE CONDUCTED AT THE U. S. ARMY ORDNANCE CENTER AND SCHOOL, DURING THE YEARS 1970 - 71.

THE AVERAGE STUDENT ENROLLMENT WAS TWELVE STUDENTS PER CLASS. ONE OF THE PREREQUISITES FOR ADMITTANCE WAS A MINIMUM ARMY STANDARD SCORE OF GM 100, WHICH WAS DERIVED FROM THE GENERAL MAINTENANCE APTITUDE AREA OF THE ARMY CLASSIFICATION BATTERY.

THE LENGTH OF THE TRAINING PERIOD WAS TEN WEEKS AND THE STUDENTS WERE REQUIRED TO COMPLETE THE COURSE WITH A 70 PERCENT AVERAGE TO GRADUATE AS A WELDER.

SEVERAL CONSECUTIVE CLASSES WERE USED TO TEST AND VALIDATE A SELF-EVALUATION INSTRUMENT. THE FIRST AND SECOND INSTRUMENTS WERE DISCARDED BECAUSE OF A CONSIDERABLE OVERLAPPING OF THE TASKS THE STUDENTS WERE REQUIRED TO CHECK WHEN MAKING AN EVALUATION OF THEIR WELDING PROJECT. THE THIRD SELF-

EVALUATION INSTRUMENT DESIGNED GREATLY REDUCED THE NUMBER OF TASKS WHICH THE STUDENT WAS REQUIRED TO CHECK, BUT THE COLUMN HEADINGS NEEDED TO BE REVISED SO THAT STUDENT EVALUATIONS WOULD BE MORE VALID AND RELIABLE. THE FOURTH INSTRUMENT WAS TESTED ON THREE CONSECUTIVE CLASSES AND PROVIDED DATA WHICH COULD BE USED FOR COMPARING A STUDENT SELF-EVALUATION SCORE WITH A GRADER'S EVALUATION SCORE FOR A COMPLETED WELDING PROJECT. THE USE AND FLEXIBILITY OF THIS INSTRUMENT WILL BE DEMONSTRATED BY SLIDE 1.

SLIDE 1 ON

I WOULD LIKE TO DIRECT YOUR ATTENTION TO THE RIGHT HAND CORNER OF THE SLIDE. NOTICE HOW THIS INSTRUMENT CAN BE USED BY BOTH A STUDENT AND A GRADER.

THE SECOND OBSERVATION WHICH I WOULD LIKE YOU TO MAKE IS TO THE LINED SPACE FOUND IN THE INTRODUCTION. THIS SPACE CAN BE USED TO WRITE IN ANY ONE OF THE EIGHT WELDS THAT ARE PERFORMED - BUTT, FILLET, ARC, ETC.

NOW LOOK AT THE TEN ITEMS LISTED UNDER THE TASKS PERFORMED. THE PERFORMANCE OF THESE TASKS ARE IDENTICAL FOR ALL EIGHT WELDS. THIS MAKES IT POSSIBLE TO USE THIS SAME INSTRUMENT FOR ALL WELDING PROJECTS.

FINALLY THE SCORING OF THIS INSTRUMENT IS A SIMPLE PROCESS OF ARITHMETIC. A RAW SCORE IS OBTAINED BY ADDING THE WEIGHTED VALUES IN EACH COLUMN. EACH X IN COLUMN 1 RECEIVES A VALUE OF FOUR, EACH X IN COLUMN 2 RECEIVES A VALUE OF THREE, EACH X IN COLUMN 3 RECEIVES A VALUE OF TWO, AND EACH X IN COLUMN 4 RECEIVES A VALUE OF ONE. WHEN THE VALUES OF THESE COLUMNS ARE TOTALED, THE SUM REPRESENTS A RAW SCORE. THE RANGE OF THE RAW SCORE IS FROM 40 TO 10. A SCALED SCORE IS OBTAINED BY DIVIDING THE RAW SCORE BY 4 AND MULTIPLYING BY 10. FOR EXAMPLE: IF A RAW SCORE TOTALED 28,

WHEN DIVIDED BY 4 AND MULTIPLIED BY 10 IT WOULD EQUAL A SCALED SCORE OF 70.

IN REFERENCE TO THE COLUMN HEADINGS A QUESTION MIGHT ARISE: HOW CAN A STUDENT DISTINGUISH BETWEEN A BETTER THAN AVERAGE, AVERAGE, BELOW AVERAGE AND A FAR BELOW AVERAGE WELDING TASK. EACH CLASS ENROLLED IS GIVEN A THOROUGH ORIENTATION WHICH INCLUDES:

(1) A TOUR OF THE CLASSROOMS AND THE WORKING STATIONS OF THE SHOP AREA.

(2) A REVIEW OF THE COURSE REQUIREMENTS WITH A EMPHASIS ON WHAT THEY MUST ACCOMPLISH TO SUCCESSFULLY COMPLETE THE COURSE.

(3) THE SHOWING OF TV TAPES WHICH COVERED THE TECHNIQUES OF THE WELDING PROJECT THEY WERE REQUIRED TO PERFORM.

(4) IF TV TAPES WERE NOT AVAILABLE PRACTICAL DEMONSTRATIONS PERFORMED BY INSTRUCTORS WERE SUBSTITUTED.

(5) IN THE SHOP AREA THERE ARE MODELS FOR ALL WELDS ON DISPLAY. THESE MODELS ILLUSTRATE THE DIFFERENCES BETWEEN WELDING PROJECTS WHEN CORRECT AND FAULTY TECHNIQUES ARE USED. THESE MODELS ARE POINTED OUT AND EXPLAINED BY THEIR INSTRUCTORS.

(6) THE STUDENTS ARE GIVEN THE OPPORTUNITY TO PERFORM PRACTICE PROJECTS BEFORE THEY ARE TESTED ON THEIR ACTUAL PERFORMANCE OF THAT PROJECT.

(7) THE STUDENTS ARE GIVEN THE OPPORTUNITY TO COMPARE AND DISCUSS THEIR PRACTICE PROJECTS WITH THOSE OF THEIR FELLOW STUDENTS.

(8) THE STUDENTS ALSO DISCUSS AMONG THEMSELVES THE QUALITY AND PROBLEMS OF THEIR PERFORMANCE DURING CLASS BREAKS.

CONSIDERING THESE FACTORS IT IS SAFE TO ASSUME THAT THE STUDENTS ARE QUALIFIED TO MAKE HONEST APPRAISALS OF THEIR PERFORMANCES ON ALL THE WELDING PROJECTS THEY ARE REQUIRED TO PERFORM.

SLIDE 1 OFF

IN THE NEXT PHASE OF THE STUDY THE SELF-EVALUATION INSTRUMENT WAS USED TO DETERMINE IF THERE WERE ANY SIGNIFICANT DIFFERENCES BETWEEN THE SCORES OF TWENTY-FIVE STUDENTS WHO MADE EVALUATIONS OF THEIR OWN WELDING PROJECTS AND THE GRADER'S EVALUATION OF THE WELDING PROJECTS OF THESE SAME STUDENTS.

SLIDE 2 ON

A COMPARISON OF THE MEAN SCORES OF STUDENT SCORED WELDS AND GRADER SCORED WELDS INDICATES THE STUDENTS SCORED THEMSELVES LOWER THAN THE GRADER. ALSO, THE SLIDE SHOWS ONLY ONE WELD IN WHICH THERE WAS A SIGNIFICANT DIFFERENCE BETWEEN THE STUDENT'S AND GRADERS EVALUATIONS.

THE DATA SHOWN ON THIS SLIDE INDICATES THAT THE STUDENTS MADE VALID APPRAISALS OF THEIR WELDING PROJECTS WHEN COMPARED TO THE GRADER'S EVALUATIONS OF THESE SAME PROJECTS.

SLIDE 2 OFF

THE THIRD PHASE OF THE STUDY IS DEVELOPED ON SLIDE 3. THIS SLIDE WILL SHOW A COMPARISON OF THIRTY-EIGHT MATCHED PAIRS OF STUDENTS. THE GRADER SCORED THE WELDING PROJECTS OF BOTH GROUPS.

SLIDE 3 ON

THIS SLIDE SHOWS THE LEVEL OF SIGNIFICANCE FOR EACH OF THE EIGHT WELDING PROJECTS. NOTICE THAT FOR THE BUTT WELD THE STUDENTS NOT MAKING SELF-EVALUATIONS WERE RATED HIGHER BY THE GRADER. FOR THE NEXT FOUR WELDS THE GRADER RATED THE STUDENTS MAKING SELF-EVALUATIONS HIGHER THAN THE STUDENTS NOT MAKING SELF-EVALUATIONS, BUT THE SCORES DO NOT SHOW ANY

SIGNIFICANT DIFFERENCES.

FOR THE LAST THREE WELDS THE STUDENTS MAKING SELF-EVALUATIONS OBTAINED GRADER'S MEAN SCORES THAT SHOW SIGNIFICANT DIFFERENCES ON THE .10 AND .05 LEVELS. THIS SLIDE INDICATES THAT WELDING STUDENTS MAKING SELF-EVALUATIONS OF THEIR PRODUCTS CONSISTENTLY SCORED HIGHER THAN STUDENTS WHO DO NOT EVALUATE THEIR WORK.

SLIDE 3 OFF

THE FOURTH PHASE OF THE SET STUDY WAS TO DETERMINE IF RETURNING THE STUDENTS' SELF-EVALUATIONS ALONG WITH THE GRADER'S EVALUATIONS AND BOTH COMPLEMENTARY AND CRITICAL COMMENTS HAD A SIGNIFICANT EFFECT ON STUDENT PERFORMANCE.

IN ORDER TO TEST FOR SIGNIFICANT DIFFERENCES THE GRADER'S SCORES FOR THESE STUDENTS WERE COMPARED WITH THE GRADER'S SCORES OF THE STUDENTS NOT MAKING SELF-EVALUATIONS.

THE NEXT SLIDE SHOWS A COMPARISON OF THIRTY-EIGHT STUDENTS NOT MAKING SELF-EVALUATIONS MATCHED WITH THIRTY-EIGHT STUDENTS WHO MADE SELF-EVALUATIONS WHICH WERE RETURNED TO THEM ALONG WITH THE GRADER'S EVALUATIONS AND COMMENTS.

SLIDE 4 ON

THIS SLIDE LISTS SIGNIFICANT DIFFERENCES IN FAVOR OF THE STUDENTS MAKING SELF-EVALUATIONS ALONG WITH THE GRADER'S EVALUATIONS AND COMMENTS RETURNED TO THEM.

NOTICE THERE ARE SIGNIFICANT DIFFERENCES RECORDED FOR SIX OF THE EIGHT WELDS. THE t FOR THE ALUMINUM WELDING PROJECT BORDERS ON BEING SIGNIFICANT AT THE .05 PERCENT LEVEL, BUT IS SIGNIFICANT AT THE .10 PERCENT LEVEL. THE DATA DEPICTED ON THIS SLIDE INDICATES THAT

RETURNING SELF-EVALUATIONS WITH THE GRADER'S EVALUATIONS AND COMMENTS MOTIVATES THE STUDENTS TO PERFORM AT A HIGHER DEGREE OF EFFICIENCY.

SLIDE 4 OFF

THE FOURTH PHASE OF THE SET STUDY WAS TO TEST FOR SIGNIFICANT DIFFERENCES BETWEEN THE GRADER'S SCORES FOR EIGHTY-ONE STUDENTS WHO MADE SELF-EVALUATIONS WHICH WERE RETURNED TO THEM WITH THE GRADER'S EVALUATIONS AND COMMENTS, MATCHED WITH EIGHTY-ONE STUDENTS WHO DID NOT MAKE SELF EVALUATIONS.

THIS PHASE OF THE STUDY WAS PERFORMED TO TEST FURTHER THE EFFECTS OF STUDENT SELF-EVALUATIONS ON STUDENT PERFORMANCES.

SLIDE 5 ON

THESE STUDENTS WERE MATCHED BY USING THEIR GM SCORES AS THE BASIS FOR FORMING MATCHED PAIRS. THE GM MEAN SCORE FOR THE STUDENTS MAKING SELF EVALUATIONS WAS 109.16 AND THE GM MEAN SCORE FOR THE STUDENTS NOT MAKING SELF-EVALUATIONS WAS 109.21 WHICH SHOWS A DIFFERENCE OF FIVE HUNDREDTHS OF A POINT. THIS IS AN INDICATION OF HOW CLOSELY THE STUDENTS WERE MATCHED.

THIS SLIDE SHOWS THAT FOR SEVEN OF THE EIGHT WELDS THERE WAS A SIGNIFICANT DIFFERENCE AT THE .01 - ONE PERCENT LEVEL. THIS INDICATES THAT THE SELF EVALUATION TECHNIQUE MOTIVATED STUDENTS TO MAINTAIN AND PERFORM AT A HIGHER DEGREE OF EFFICIENCY THAN THOSE STUDENTS WHO WERE NOT REQUIRED TO MAKE SELF-EVALUATIONS. NOTICE THAT THE MEAN SCORES FOR THE BUTT WELD ARE COMPARATIVELY EQUAL AND THERE IS NO SIGNIFICANT DIFFERENCE BETWEEN THE MEANS. I WOULD ALSO LIKE TO DIRECT YOUR ATTENTION TO THE CONSISTENCY OF BOTH SETS OF SCORES. THE STUDENTS WHO DID NOT MAKE SELF-EVALUATIONS

SCORED CONSISTENTLY AROUND A GRADER'S MEAN OF ABOUT 72 WHILE THE STUDENTS WHO MADE SELF-EVALUATIONS SCORED AT A MEAN OF ABOUT 75, WHICH IS AN AVERAGE OF ABOUT THREE POINTS HIGHER. THE DATA SHOWN ON THIS SLIDE PROVES THAT THE SELF-EVALUATION TECHNIQUE WORKED AS A MOTIVATING FACTOR BECAUSE OF THE HIGHER PERFORMANCE SCORES OBTAINED BY THE 81 STUDENTS WHO MADE SELF-EVALUATIONS AND HAD THEM RETURNED, ALONG WITH THE GRADER'S EVALUATIONS AND COMMENTARY FEEDBACK.

SLIDE 5 OFF

FINDINGS OF THE SET STUDY

THE FINDINGS OF THE SET STUDY CAN BE SUMMARIZED BY THE THREE STATEMENTS WHICH FOLLOW:

(1) THIS STUDY PROVED THAT WELDING STUDENTS CAN EVALUATE THEIR PERFORMANCES COMPARABLE TO THE EVALUATION OF A TECHNICALLY QUALIFIED GRADER.

(2) THIS STUDY PROVED THAT STUDENTS MAKING SELF-EVALUATIONS SCORED PROGRESSIVELY HIGHER THAN STUDENTS NOT MAKING SELF-EVALUATIONS.

(3) STUDENTS MAKING SELF-EVALUATIONS WHICH WERE RETURNED TO THEM WITH THE GRADER'S EVALUATIONS AND COMMENTARY REMARKS SCORED SIGNIFICANTLY HIGHER THAN STUDENTS NOT MAKING SELF-EVALUATIONS.

INCIDENTAL FINDINGS

SEVERAL INCIDENTAL FINDINGS THAT FURNISHED ADDITIONAL INFORMATION WERE DISCOVERED WHILE THE SET STUDY WAS IN PROGRESS. THESE FINDINGS ARE AS FOLLOWS:

(1) UNDECTECTED PERSONAL PROBLEMS AFFECT STUDENT PERFORMANCE.

(2) STUDENTS WITH LOW LEVELS OF SELF CONFIDENCE ARE EASILY IDENTIFIED.

(3) EXTRANEOUS DUTIES ASSIGNED TO STUDENTS AFFECT STUDENT PERFORMANCE.

(4) STUDENTS WITH LOWER GM SCORES TEND TO IMPROVE THEIR SELF-EVALUATIONS AS THEY PROGRESS THROUGH THE WELDING COURSE.

(5) STUDENTS WITH HIGHER GM SCORES TEND TO MAKE MORE VALID SELF-EVALUATIONS.

(6) STUDENT SELF-EVALUATION SCORES TEND TO BE LOWER THAN GRADER EVALUATION SCORES.

(7) STUDENTS WHO NEED COUNSELING ARE EASILY IDENTIFIED AT THE BEGINNING OF THE COURSE.

CONCLUSION

THE FINAL CONCLUSION REACHED FROM AN ANALYSIS OF THE DATA OBTAINED FROM THE SET STUDY CAN BE STATED AS FOLLOWS:

STUDENTS MAKING SELF-EVALUATIONS WHICH ARE RETURNED TO THEM ALONG WITH THE GRADER'S EVALUATIONS AND COMMENTARY REMARKS SCORE SIGNIFICANTLY HIGHER ON THEIR WELDING PROJECTS THAN STUDENTS WHO DO NOT MAKE SELF-EVALUATIONS.

THE NEED FOR NEW DEGREE-AWARDING METHODS

by
Mr. Jack N. Arbolino
Executive Director
Council on College-Level Examinations
College Entrance Examination Board

It is my recommendation that to expand nontraditional programs and to increase credit by examination, to recognize the tremendous educational programs of the military, and to motivate individuals to engage in independent study, America must establish a federally-chartered national university.

Of course, we must make some conceptual changes regarding time and money as they relate to the degree. And we must substantially increase the granting of credit by examination in our regular degree programs. If we do these things, we do a great deal, but I think something more remains. We must, in addition, establish a federally-chartered national university which will award external degrees and, with those colleges which elect to participate, also grant joint external degrees for which part of the work is done in the participating college and part completed by examination through this so-called national university.

If this recommendation seems excessive, I can only assert my belief that we should reach for major institutional reform, one that will enable us to provide humanistic and social incentive to match our technological advances, and one that will enable us to honor achievement no matter where or when it occurred. If our reach does not exceed our grasp we will not only close some gaps but we open a new way to the development of individual potential and deliver at last the quality of opportunity that we have always promised. Columbia's new president, William J. McGill, says, "We form in as large institutions as possible only when people are running scared. Believe me, in higher education", Dr. McGill says, we are scared. And thus the next decade is likely to produce reorganization, curriculum reform, redefinition

of professional life, and a variety of innovations unlike anything seen in the last fifty years. Our survival depends upon it."

Well, if Dr. McGill is right, and I think he is, then there never was a more propitious time than the present for reform and renewal in higher education. That may be optimism of a very low order, but the older I get the more I realize the smallest advantages are not to be despised.

I'm not going to pose as one who knows something about degree requirements. The truth is I don't understand them. Liberal arts are hard enough to describe. They are what you need so that when you knock on yourself, somebody is home. Or they are what do not enable you to make money. Or, they are circumscribed by Jacques Barzun's aphoristic wisdom that "That liberal arts can be taught mechanically, and mechanics liberally." These thoughts, it is true, do not lead to a clear and easy definition of the liberal arts, but they do make sense. Our degree requirements, on the other hand, very often seem not to do this. They are like arbitrary curatives. They seem to say to students, "I may seem strange to you, but that is because you have defects you do not recognize. I will erase them." Within the profession, they seem to elicit a tacit exchange of tolerance. "We have a few quaint ones and so do you, but after all, we're all in this thing together." Only a few things about them are certain. They cannot be ignored, and they are usually presented with pride... "The founders of Halcyon College, valuing clear speech and serenity," and we read on to see that "calm composition" is required of all freshmen.

What does a degree mean by itself when compared with another? Does it mean the holder has racked up 120 points? Does it represent an acquaintance with the major branches of knowledge, or the ability to move with understanding within one? Is it a state of being, to be forever lighted by the glow of learning, or is it the ability to swim 100 yards? Is it all or any of these things? And what are the things a student must have in order to see himself through? Not the courses he must pass, but the qualities he must display.... patience, time, a fairly generous father, money, stamina-- these are degree requirements, too. To these requirements concessions are rare, and the admonitions are clear. Stay the course, serve your time, pay your fees, stay in line. To these requirements our institutions seldom

yield, and ironically, it is perhaps just to these where most we should. If without fear or favor, we were to examine our higher educational system, and ask Harold Howes question, "Do our institutions meet the needs of individuals, or is it the reverse?" we might be forced to admit that, though we do bend a good deal in the matter of course requirements, when it comes to other kinds of requirements we are rigid unto death. French IV can be cleared, but time and money waived for no man.

I remember with what expansiveness a committee on instruction on which I served at Columbia declared the foreign language requirement fulfilled for an American Indian who claimed he knew the literature of the Sioux. And how was this ascertained, for although the university taught more than 50 languages, we had no professor of Sioux. Ah, but there was a way. An anthropologist on the committee, one who did more than just study man, sent the student to a museum and somehow--perhaps by tomtom--word got back. "He qualifies." There were no questions. No one asked how do we know. Is there a Sioux literature? How is it recorded? In smoke? On the wind? Is it foreign to him? There were no questions because the committee wanted to do it, and I suppose we were right. But, what if he asked for a waiver of points? or dollars? Where would our expansiveness have been then?

If we really want to close the gap between individual needs and degree requirements, if we really want to expand opportunity, and encourage self-reliance, we should recognize that degree requirements consist of more than courses and subjects, and we should recognize too that degree requirements is a euphemism for institutional needs. The residence requirement, for instance, is of far more interest to the bursar than to the director of residence.

Perhaps for discussion a useful antithesis to consider is individual needs versus institutional needs, and to say institutional needs take precedence over individual needs may be like saying we love Catholicism and everything about it except Catholics.

I presume those in charge -- and I'll admit they aren't always easy to identify, and once identified they certainly aren't easy to move -- do, I think, want to improve American education. And I presume we really do

want to serve individual needs without weakening institutional needs or our institutions. I think we can better serve individual needs and at the same time improve higher education and strengthen our institutions. Moreover, we create a false antithesis when we put the needs of the individual in opposition to the needs of the institution. We have, I think, often in dead seriousness, tended to make just that dangerous antithesis. If one is helped, the other is hurt. If the student is happy, the institution must be sad. The simplifier loves the simpler single way. And we are forever confronted with alternatives in education. If you are for small colleges, you've got to be against large ones. If you believe in community colleges, how can you countenance the Ivy League? If you are forever grateful for a teacher like Mark Van Doren, how can you believe in credit by examination? Well, it's easy. Just think of some of the other teachers you had. And it's even easier if we remember what Henry Adams says in his education: "Simplicity is the most deceitful mistress that ever betrayed Man."

I said at the outset that if we really want to close gaps between individual needs and degree requirements we would have to (a) make conceptual changes regarding time and money as they relate to the degree; (b) substantially increase the granting of credit by examination; and (c) establish a federally-chartered national university which will award external degrees. We already possess the key to a b, and c. It is the College-Level Examination Program.

CLEP is a young college board program which, with the generous support of the Carnegie Corporation of New York is starting to make an impact on higher education. Let me cite just a few institutional uses of the program as examples. The University of Iowa authorizes CLEP general examinations as an alternate way to meet degree requirements. Between the fall of 1966 and December 1969, 1235 out of 1531 matriculated students who took the tests did well enough on one or more to receive a total of 5,124 hours of credit. The University of Nebraska at Omaha last year graduated 800 students who averaged 20 semester hours of credit by examination on the basis of CLEP tests. The University of Texas, having announced that the CLEP American Government subject examination would be accepted in lieu of the required courses, found itself with a small riot on its hands when it ran out of test tickets one day last November. More than 5,000 students took the test last fall. Now the University is contemplating the same application of the American History Test. The saving in instructional costs alone was $1.6 million in one year at the University of Texas.

I could cite many, many more examples of how this program is starting to make a tremendous impact--let me do just two more. The California State Colleges, under Chancellor Dumke, have decided that San Franciso State this fall will administer the CLEP examinations to all its entering freshmen. If they do well enough, they will begin as sophomores. Imagine the saving to that. If this works at San Francisco State -- and Hayakawa the President and Dumke have written a letter to every freshmen entering the state colleges -- if this works it will be adopted at all 19 state colleges next year. That will have tremendous impact on American education.

The University of Maine has just given the general examinations to 650 of its 800 entering freshmen, and to me one of the most interesting things is that the University of Maine Academic Vice-President wrote to the president of San Francisco State and said, "We hear you're doing this. Tell us about it. We're just starting." And the man at San Francisco State wrote back and said, "Thank God, we've found someone else who's doing it." Now the job that some of us have, including Ted O'Connor, is to fill in the whole country between California and Maine doing this kind of thing.

Through the national program by which the public at large is served (not the institutional program I have just described), examinations are available to unaffiliated students once a month at 60 centers located in large cities throughout the country. At the moment, almost 1000 colleges will give credit on the basis of the CLEP tests, and although the program is not yet a pervasive force in higher education it has grown steadily as its use has become understood.

I'll stop with that description, but the important point is that CLEP is American higher education's most exciting way to help colleges grant placement and credit by examination. It is also the key to the flexibility we must have if we want properly to recognize all the proper, appropriate, and superior work that goes on in the Armed Forces. The largest batch of takers in the CLEP program are the ones who have taken the examinations through USAFI. And if we want to expand nontraditional learning, this is the program to use. We can help students and help ourselves, too, but to do both we must make concessions. We must recognize that there are other ways to learn other ways to measure, other ways to keep time in academia.

There are serious questions. For instance, is an external degree a threat to traditional institutions of higher education? What is the extent and kind of interest and receptivity to an external degree that exists among the public at large, and among our college and universities? Should the effort serve adults exclusively, or should the concept of continuing education be interpreted to include the traditional college age group? There are no end of questions. One I think I would like to cite is to what extent can the considerable educational work going on outside our colleges, in the military, in business, industry, and government be recognized toward a degree? Can such a recognition be effected other than by examination?

Well, we had better recognize some of these new ways, for if we persist in yielding only to those student needs which are easy to meet and continue to deny those which force us to think and change a little we could end up as lonely caretakers polishing cannons in the public square, or maybe even worse--brass on a sunken ship.

I think we are starting to sound a little hollow right now in American education. How much longer can a faltering system, sanctified by custom, besieged by shock, and led by less than supermen continue? Maybe forever, but I don't think so. Certainly, institutions die hard and maybe it is true that men will change their morals before they dare to change their institutions. However, if we do want to stand still, no recent study or pronouncement by any responsible group or individual gives us much cause for comfort. And even if we see education as a vast preserve, richly stocked with fat slow-moving targets, we certainly cannot regard all who warn us as professional alarmists.

I think it is clear that the wind is rustling and we cannot stand still. Perhaps we need not scramble, but at least we'd better stir ourselves a little. John Gardner says, (another Marine, if you'll pardon the expression) "Passivity is curable." If that diagnosis is not correct, we might as well give up right now. Frequently hope may be fallacious but invariably despair is deadly.

How then do we move to close the gap? How do we move better to serve students and institutions? How, if you like, do we strengthen our country? We can

begin by taking things out of opposition by acknowledging that there are several kinds of degree requirements, and by accepting credit by examination as the key to flexibility for individuals and institutions. The working system of credit by examination within a college can open the door to functional, not paper, programs of independent study, and to the conservation of time and resources by both students and colleges. It can do more. It can lay the groundwork for the growth of institutional external degree programs that will enable each college to participate in a movement which will truly open education to all our citizens.

A national university, perhaps federally chartered, which would award external degrees, would be a sound supplement to the cause of higher education and could alleviate some of the pressures on our established institutions. Such university, if it did not engage in instruction and incidentally I do not think it should, for we have many sources of instruction which could absolutely, properly, and appropriately be measured and meshed with such a university. I do not think this university should engage in instruction, and it might seem to compete with and threaten our established institutions, and maybe it would. Perhaps the answer to that threat, if it is real, lies in the establishment of joint external degrees. Degrees in which some of the requirements are met through the national university and some through those colleges which choose to participate in a joint program with the national university.

Of the three kinds of external degrees just mentioned, the institutional, the national, and the joint external degree, the first two have obvious weaknesses. If all we had were the institutional external degree with, let us say, two or three hundred colleges offering them, we would risk proliferation, confusion, fragmented effort, and duplication of validating instruments. If all we had were the national external degree, that is, the one federally-chartered national university, we would risk the threat previously mentioned and the strictures of the monolith. The third, the joint external degree, depends on the establishment of a national university, and together the last two approaches might constitute the best plan. This could give us a new institution--a national university which would award degrees by examination, and a joint program in which colleges that wish to participate would, with the national university, award degrees completed in part by examination. The participating colleges would, of course, continue to award regular degrees as well.

I began this look into the future by saying credit by examination is the key to flexibility for individuals as well as institutions. I would summarize by saying that I think this flexibility is essential, that we ought to reconcile individual and institutional needs. We have been willing to tinker with course requirements. We must do far more than that with the other requirements that bind us. We need to do more than just wait for things to change us. I think it is clear that we must advocate change and engage in self-renewal.

We can face some wobbly times. Broken patterns will be commonplace and so will troublesome new methods and measures. Something else may happen, too. There may be a true reconciliation between scholarly and societal goals, and a rapprochement at least between our formal system and some of the learning that goes on outside it. That really would be something. We can confront and affect these changes together and maybe we had better, for it seems dubious that separate uncoordinated institutional attempts will get us through intact.

Finally, I think we cannot and should not avoid some form of the external degree. A national university with provisions for joint awards with participating colleges seems best to be. If we shape this part of our future and do not let it shape us, we may one day see as providential the problems that afflict us now.

ANNUAL CONFERENCE OF THE UNITED STATES TESTING ASSOCIATION - WASHINGTON SEPTEMBER 1971

PREFERENCE CONSISTENCY TESTING: FLEXIBLE VALUE SYSTEMS AND ASSESSMENT RELIABILITY

SQN LDR B N PURRY MA RAF

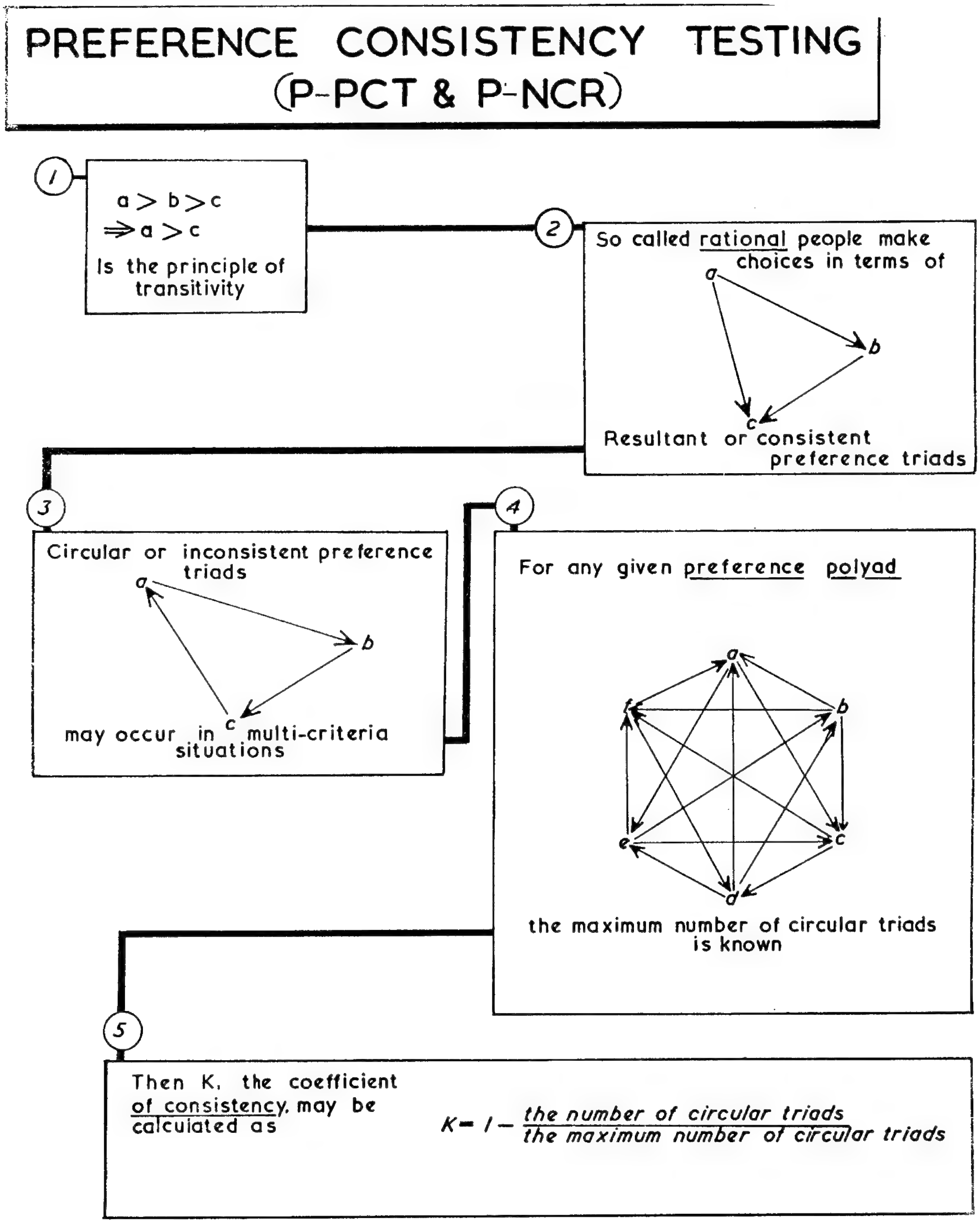
INTRODUCTION

1. King Alfred who ruled the Kingdom of Wessex from AD 871 to AD 901 is reported in a law book written in the 14th Century attributed to Andrew Horn (6) to have been particularly displeased one year with the conduct of his judges - and in that year he gave orders that no fewer than 44 of them should be hanged. For some time now I have been working with various people who are required to judge human behaviour and my sympathies lie with King Alfred!

2. There is a requirement to talk with various groups about problems of assessment. The aim of my talks with them is not so much familiarisation with the mechanics of our reporting systems (these they should already know) but rather the indication of some of the dangers and pitfalls they may expect to face in their use.

3. You will not be surprised to hear that much of the time is taken up with explanation and discussion of the basic concepts of variance, reliability, and validity. Nor will you be surprised when I tell you that one leans heavily on American research for examples and analyses of some of the more common difficulties found in interpersonal perception and inherent in the rating of unobservable qualities of the human mind. Lt Col Glory Sturiale's delightful papers on inflation, and her work on the predictability of patterns in "officer effectiveness reports" (12)(13) provide excellent illustrative material for the discussion of halo and for the introduction of ideas of self projection and personal constructs. McKendry and Lindsay's findings at Lackland that "raters favour personality trait check lists" (8) are incorporated, whilst the fundamental issue of objectivity

FIGURE AND SLIDE No 1



- the issue of clinical versus statistical prediction is discussed by reference to the authoratative reviews of Meehl (9) and of Sawyer (11). And one touches on the well established inadequacies of brief interviews as described so succinctly by Lt Col Cary Thompson Jnr (14).

4. To bring matters nearer home one relies on data collected and published in the Royal Air Force but not widely distributed. There are reports on the workings of the assessment system in use at our Officer Cadet Training Unit (1)(15) and there are a number of other research studies and papers concerning human performance in the services on which one can draw but despite the volume and variety of this evidence none of it is quite as persuasive as one would like and none of it one feels actually induces a permanent attitude change. One fears the students really think as a veteran of the English bar once put it, "there may have been bad judges, and there will be bad judges, but there are no bad judges now" (16). My problem - since I have neither the authority nor the power of King Alfred - though sometimes I wish I had - is to persuade the groups I work with that there are difficulties and at the same time to try to resolve some of them by an analysis of their occurrence. With each trainee group therefore, and to these ends, I play, what I call "games". (some people kindly call them experiments, whilst some even go so far as to describe the whole thing as "research"). It is one little set of these games that I would like to describe to you this morning.

THEORY, AND FIRST HYPOTHESIS

5. Forgive me if I refresh your memory with some very elementary theory (OHP Slide 1) In mathematics and in the physical sciences most operators have the property of transitivity (Figure 1 Block 1). If A is greater than B, and B is greater than C, then this implies that A is greater than C. And we may draw this relationship like a triangle of forces and then call it a resultant triangle. If this graphic is used to depict a simple system of human choices then - since it is a rational

way of decision making - we describe it as a consistent triad. (Figure 1 Block 2) But human choices are not all always made in this way. I may for example prefer apples to oranges because I prefer their taste. I may prefer oranges to bananas because the colour is so much richer and more attractive but given the choice of apples and bananas I may prefer bananas because they peel so much more easily. This set of decisions could be drawn as a circular triad and it would be described as inconsistent because the criteria of judgement were changed. (Figure 1 Block 3) Given a set of interlocking paired comparisons in which each member is compared with every other it is possible to construct a preference polyad (Figure 1 Block 4) and to express an assessor's consistency numerically in terms of the extent to which he achieves all consistent triads. (His score is actually one minus the ratio of circular triads to the maximum circular triads possible, and hence K, the co-efficient, varies between zero and one.)

6. The numeracy of this technique is well described in elementary textbooks (5) (10) and the fact that this inconsistency or irrationality in human judgements exists was again commented upon recently by professor Jaspars in a paper given at this year's NATO Defence Psychologists Symposium at Brussels (7).

7. It seems therefore that one might reasonably consider three questions:

a. How many assessors are inconsistent in this sense?

b. Can inconsistency be measured as easily as the theory implies?

c. Is there any connection between this inconsistency and the low reliability found with monotonous ease and regularity between assessors required as another exercise to assess characters in a film?

METHOD AND RESULTS 1

8. Subjects were each asked to imagine themselves to be the owner of a night club in need of a new receptionist. It was assumed that in making a selection from

a number of applicants, academic qualifications and intellectual prowess could be disregarded in this situation! (35 mm Slide 1) The choice was to be made on appearance. Subjects were then presented with a succession of paired pictures of the competing girls with instructions to record their preference from each pair. This now internationally famous, "Purry nightclub receptionist test" (35 mm Slide 2) or P-NCR test included 2 sets of 6 girls, each girl in a set being compared with every other girl in her set, and precautions were taken to prevent "cheating" or "bias". Five additional pairs were put into the combined packs of 2 sets as distractors and no girl was consistently placed on the left or on the right (Slide Off).

9. Whilst not surprisingly the night club receptionist game is regarded by young male subjects as good fun, those found to be inconsistent in their preferences express real concern at the discovery of an irrationality they previously thought impossible. Indeed- since the very term "inconsistent" seems to have a pejorative connotation and it would be quite wrong to imply any value judgement from these experiments, the point is explicitly made that the test may in fact be distinguishing between those who are rigid, hide-bound and mentally moribund on the one hand, and those who have a laudible flexibility on the other!

10. In the event P-NCR 1 was administered to 114 subjects of whom about half were found inconsistent (or flexible) by the method described which confirmed that inconsistency in this sense could easily be identified. The exercise assessments given by these subjects to characters in a film were then sorted into two groups (those given by consistent subjects and those given by inconsistent subjects as now measured by the P-NCR test). Co-efficients of concordance were calculated for each group and there was effectively no difference found between them. Thus the hypothesis of readily recognisable correlation between consistency and reliability had to be rejected.

(L) A	Has a satisfactory trade knowledge and achieves a satisfactory standard in his current duties.	(R) A	Possesses a very good trade knowledge is extremely proficient in his current duties and can cope with unusual problems.
	He is always in the lead when the situation demands and is outstandingly effective in control Organises and plans efficiently.		He is always in the lead when the situation demands and is outstandingly effective in control. Organises and plans efficiently.
	Exceptionally quick to get to the root of a problem and is sound reasoned and consistent in judgement.		Exceptionally quick to get to the root of a problem and is sound, reasoned and consistent in judgement.
	He is highly and justifiably self-confident.		He is highly and justifiably self-confident.
B	He is completely reliable thoroughly conscientious and outstandingly eager to serve and enthusiastic in fulfilling both primary and secondary duties.	B	He can be relied on to complete all normal duties satisfactorily. He fulfils his normal service obligations.
	He is outstandingly smart at all times and presents an absolutely first rate appearance.		He is outstandingly smart at all times and presents an absolutely first rate appearance.

OF THESE TWO ON PAGE 17

I WOULD PREFER TO PROMOTE (________)

OHP SLIDE AND FIGURE 2

ATTRIBUTE PREFERENCE AND CONSISTENCY

11. The method was then applied to a situation with more apparent relevance to the service situation Subjects were to imagine themselves on an NCO promotion board at which for convenience word pictures of the candidates for promotion were presented in a series of paired comparisons. There are in fact 6 candidates in this test (although subjects are not told this) and as in P-NCR each is compared with every other - and a number of distractors are added. Each candidate is described in terms of 6 attributes five of which he possesses in an extremely high degree and one merely at a satisfactory level (OHP slide and Figure 2). Thus each paired comparison requires the subje ct to choose between candidates having 4 identical attributes and differing only in the remaining two. One candidate will have outstanding quality A with satisfactory B whilst the other has satisfactory A with outstanding B and each subject's forced choice is in fact between attribute A and attribute B. Hence the title of the test 'Attribute Preference and Consistency Testing' or for short 'APAC'.

APAC RESULTS

12. As with P-NCR, APAC testing finds about half of subjects to be inconsistent but because of its face validity, subjects found inconsistent become even more thoughtful about the problems of assessing than those found inconsistent by P-NCR.

13. The second immediate benefit in the small group training situation is that the preference hierarchy of consistent subjects can very quickly be drawn up on a chalk board to demonstrate the fact (which has not yet failed to manifest itself) that though each consistent subject obviously knows which of the military attributes he has just been choosing between he prefers, and has a rank order of those preferences, when one compares the rank orderings of the several consistent subjects there is very little agreement between them. One is reminded of Lord Dunedin's somewhat cryptic aphorism "no judge is always right, but Judge Lindley even if he were wrong, was consistently sound!" (3).

14. A second, and more lengthy, P-NCR test was designed and administered to subjects who also worked the APAC test. Though both tests discriminated between consistent and inconsistent subjects there was no correlation between them. It was concluded that whatever the test measured it was not an enduring disposition operating in all circumstances and at all times. Inconsistency it seems is specific to situations.

15. The mean attribute preference hierarchy of inconsistent subjects in the APAC situation was compared with the mean profile of consistent subjects and no significant difference was found between those mean, but the variances of ratings about the mean were found to be significantly different. Inconsistent subjects have low variance: consistent subjects, high: which seems to make sense. The subject whose preferences in a given situation are so indecisive that he is actually inconsistent in his choices is surely the subject who dithers about a middle of the road position - in other words he is sitting right on the fence. Conversely, the consistent subject is one whose mind is made up (for better or for worse!) and he expresses his view with determination. Thus a confidence or certainty factor seems to be involved and this is now the subject of further enquiry.

CONCLUSIONS

16. One is conscious that information processing is a complex subject and considerable methodological sophistication is necessary if we are ever to understand it. One has only to read the work of Cohen (2) on the processing of contradictory information, of Guildford (4) on the methods of paired comparisons of Jaspers (7), and of Warr and Knapper (17) on person perception, to realise the extent of the problems in this field and so I hesitate to suggest conclusions based on the little studies - the games - I have talked about this morning. But some conclusions are it seems, there to be drawn.

17. Inconsistency is an established fact but it is not an enduring trait to be found in some people and not in others. It seems to be a characteristic of

performance by some people in certain situations. Given a specific assessment situation - a selection centre springs readily to mind - it should be possible to design a specific and simple test of consistency in that situation and the test itself may be used to provide the basis for improving performance by training. Assessors found to be inconsistent in a given situation should not be employed in it. There seems little point in using a human judge whose criteria are demonstrably irrational and unpredictable, but the fact that one may demonstrate the irrationality - that one may pinpoint quite precisely the exact nature of the error-leads one to hope that corrective action may be taken.

18. The judgements made by consistent appraisers are capable of being monitored - and should be. Further, since there is a clear link between inconsistency and the error of central tendency, the judgements of consistent appraisers should always be tempered with the independent judgements of others.

19. In these ways we may perhaps reduce the numbers of judges who should be hanged!.

REFERENCES

(1)	Adamson A:	RAF OCTU assessment and reporting system - Report of Working Party - 1967.
(2)	Cohen R:	Processing contradictory information about other people - 1971.
(3)	Dunedin, Lord:	The Times 18 Feb 1932.
(4)	Guildford J P:	Psychometric methods 1954.
(5)	Guildford J P:	Fundamental Statistics in Psychology and Education.
(6)	Horn A:	The Mirror of Justices - 1320 (SIC).
(7)	Jaspars J M R:	Person Perception and Social Cognition - A Review and Discussion of Major Problems - 1971.

(8)	McKendry & Lindsay:	A Word Picture Check List for Officer Effectiveness Report - 1964.
(9)	Meehl:	Clinical versus Statistical Prediction - 1954.
(10)	Moroney:	Facts from Figures.
(11)	Sawyer:	Measurement and Prediction, Clinical and Statistical - 1966.
(12)	Sturiale G:	Officer Performance Evaluation - A System within a System - 1971.
(13)	Sturiale G:	Predictable Patterns in Air Force Effectiveness Report Factor Ratings - 1969.
(14)	Thompson C A:	Predicting Performance at the Air Force Academy from Interviews - 1962.
(15)	Tilley K W:	Officer Quality Assessments at OCTU - 1967.
(16)	F J de Verteuil:	50 Wasted Years.
(17)	Warr & Knapper:	Perception of Persons and Events.

USING SELECTION DATA FOR LONG TERM EVALUATION AND PLANNING*

by ARTHUR GARDNER

Senior Psychologist (Naval) Division
Department of the Chief of Naval Research
Ministry of Defence (Navy)
London, England

The opinions expressed are entirely those of the author and do not necessarily accord with those of the Ministry of Defence (Navy).

* Prepared for presentation at the 13th Annual Military Testing Association Conference on "Improved Techniques and Procedures in Personnel Evaluation", Washington DC. 20-24 September 1971

USING SELECTION DATA FOR LONG TERM PERSONNEL EVALUATION AND PLANNING

BY ARTHUR GARDNER

Senior Psychologist (Naval) Division
Department of the Chief of Naval Research
Ministry of Defence (Navy)
London SW1
England

As Sands (1970) pointed out to last year's MTA Conference, there is a growing body of opinion which is dissatisfied with the traditional correlational approach to the evaluation of selection systems. The argument is in two parts: firstly that a correlation coefficient is not necessarily the most appropriate measure of validity; and, secondly, that, even if it were, good selection requires more than just the identification of valid instruments - it requires also the identification of those policies which will make the best use of those instruments. To these ends many psychologists would probably agree with Sands in advocating a cost-effectiveness approach as a more useful alternative.

Our own experience in the Royal Navy provides, I think, some insights as to why this has arisen. Psychologists first became involved with large scale selection in the British Armed Forces during the 1939-'45 War. Their problem then was to sort a wide variety of Conscripts (ie Draftees) into specific jobs for a fairly limited duration. Their solution was the expedient one of operating cut-offs on predictors validated, of necessity, against short term training criteria. Justified though this may have been it has influenced our thinking, I believe, for too long. Consider, for example, some of its basic attributes:

(1) The use of cut-offs requires a large and continuing supply of capable Recruits.

(2) The use of cut-offs implies the concept of "substitutability" (ie that any 100 men, say, above the cut-off are as acceptable as any other 100 men also scoring above the cut-off).

(3) The use of cut-offs, unless used explicitly as quantity/quality regulators, assumes the existence of all-or-nothing relationships between predictors and criteria (ie that no one succeeds below the cut-off).

(4) As mentioned earlier the whole system was evolved using short term criteria.

Now, how appropriate is such a system for the present day? Arguably, two broad changes have occurred. The most important, perhaps, is that in Britain we no longer use conscription (the draft) as a means of recruiting. In consequence, the Royal Navy is no longer assured of a ready-made supply of

highly qualified Applicants. In fact, the Navy was recently facing quite severe manpower shortages! Furthermore, these shortages affected both quantity and quality which, as a result, drew attention to the fact that men do differ with respect to their potential for attaining organisational goals - even amongst those with scores above the old established cut-offs! In other words, the assumptions of stable inputs and substitutability were seen to be increasingly inappropriate. Incidentally, they further lost credence as a result of the increasing differentiation between Branches (eg between the Seaman Branch and the technical Branches) in terms of the work done and their perceived attractiveness to Applicants.

The other main development affected the methodologies of personnel selection and manpower planning. The trend away from traditional approaches to selection research and towards a cost-effectiveness approach reflected a growing concern with systems thinking. By this I mean the tendency to see personnel selection as but one component of a larger manpower system. In the early days there was little need for this, nowadays, especially in a volunteer Service, it is essential. One sees the same trend in manpower planning. That is, a growing concern with career planning and the long term. I do not mean to suggest, of course, that this is completely new. For many years, for instance, the Navy calculated a Petty Officer Potential (POP) for each Branch as an aid to long term planning. The POP, however, was simply the number of men with scores above a certain level. A later refinement changed this to "the number of men with scores above a certain level divided by 4" to correct for low engagement rates, etc. Clearly, however, such a forecast is too simple-minded to be entirely satisfactory, especially when one realises that some Petty Officers had scores well below the POP "cut-off".

This brings me to the question of the nature of the relationship between predictors and criteria. To my knowledge we have never identified a "real" all-or-nothing cut-off. All that has been found is that sometimes high scorers do better than low scorers but that some low scorers do as well as some high scorers. In other words, we have never identified an absolute qualitative basis for employing rigid selection cut-offs. What is known, however, is that, generally speaking, low scorers tend to "cost" more, as a group, than high scorers (eg by virtue of higher wastage rates or slower progress through training). Nevertheless, there is no evidence to date - to my knowledge - which suggests that no low scorers are worth employing (cf the US experience with "Project 100,000").

Thus, I would argue, that as a result of many separate but inter-related changes the traditional approaches to selection (and allocation) have been overtaken. They had within them, however, the potential for further development and it is this which I want to consider in the remainder of the paper.

In particular it has been found that the data collected for selection purposes can be used to evaluate the future of the system for which the men have been selected. And it is becoming possible to make a useful contribution to the assessment of the effects of quality as well as quantity on the future of the organisation. For example, it is possible to say not only how many men are required to maintain some state of the system but also to revise this estimate in the light of personnel measures relating to the quality of those men. Further, it is possible to envisage quantitative requirements being changed or indeed partly determined by the measured quality of the inflow.

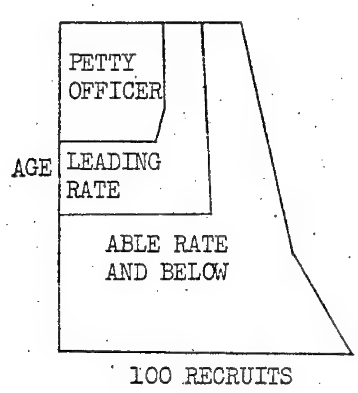
Perhaps this may be best described as an actuarial approach to manpower requirements - certainly it is increasingly the case that we are able to attribute varying degrees of risk to cohorts of entrants who previously would have been regarded as homogeneous in the sense of having been assessed as 'suitable' for the training and jobs the organisation required them to do.

Some examples will, I hope, make this clearer.

Manpower Planning in the Royal Navy

A principal concern, increasingly, is to relate personnel selection policies more closely to manpower planning objectives. In the Royal Navy Manpower planning is based upon steady state modelling (cf Jones 1969). Put simply, this amounts to specifications of size, structure and preferred quantitative development of the organisation.

Figure 1 - Hypothetical career model for Naval Ratings



In Figure 1* the bottom line represents an arbitrary 100 recruits and the other lines show how these diminish or are advanced in Rate as the years go by. Looked at vertically the diagram shows the composition of the Rates that would result eventually if 100 Recruits joined each year and wastage remained as assumed (ie it shows the steady state).

Management's task, once the steady state has been defined, is to ensure that it is achieved. In practice this involves action to avoid potential imbalances between what is predicted and what is desired. Personnel selection policies, insofar as they affect the quality and quantity of the input to the manpower system, provide an early opportunity for doing this (ie for engineering the eventual composition of the organisation).

In passing, I would point out that the selection stage is not the first such opportunity. In a volunteer service at least, the initial problems are to attract Enquirers and then to turn these into Applicants - only then can selection proper begin.

But to concentrate on the selection stage, the policy problem is one of optimising the choice of Applicants from the point of view of the overall system objectives. To this end you have to think about more than just the fitness of a particular individual for a particular job. You have to think about late criteria (eg advancement to higher Rates) as well as immediate training criteria. You have to try to recruit enough people at given ability levels to ensure an adequate supply of future leaders, etc. In other words, you have to try to maintain the steady state illustrated by Figure 1. One way of doing this is by using actuarial tables based on selection data (cf Vernon 1964).

Actuarial Tables in Selection

Table 1 shows the substantive advancement of Seamen, who entered the Royal Navy between 1946 and 1950 and were still serving in 1956, in terms of their T2 grades. (The T2 test battery is factorially similar to the AFQT but the grades, as measured against a standard Applicant population, are: A - top 10%; B - next 20%; C - middle 40%; D - next 20%; E - bottom 10%)

---

* What Jones (1969) called an 'actuarial' or 'service' table

Table 1 - substantive advancement of Seamen

		% serving at each Rate			
		Able	Leading	PO +	
T2 Grades	A	39	37	24	100%
	B	66	25	9	100%
	C	80	16	4	100%
	D	91	8	1	100%
	E	93	7	-	100%

(from SP Report QC 33)

Actuarial or expectancy tables of this type have, of course, been used for many years. Nevertheless, they have an advantage over many methods in being directly relevant to steady state planning. For instance, Table 1 shows that a given number of POs can be produced in a variety of ways dependent upon the T2 grades of the Entrant cohort (eg note that relatively more men of grade B must be recruited in order to produce the PO quota than would be necessary were they all of grade A*).

Obviously T2 is a potentially useful predictor but, as is well known, the prediction can be improved, at least in some cases, by a more multivariate approach. Tables 2-4 illustrate this in terms of two-way actuarial tables (ie Recruiting Test Scores and Educational Attainment considered in conjunction). Clearly the segmenting could be made even more complex and comprehensive.

---

* One must assume that in the absence of men of grade A, say, the proportion of grade B men advanced to PO would be as shown in the table.

Table 2 - Seaman still serving after 6 years

Educational Attainment on Entry **	Recruiting Test Scores	Proportion Still Serving	N
With GCE	High	.90	50
Without GCE	High	.82	50
"	Medium	.75	49
"	Low	.85	54

(from Barnecutt, 1968)

Table 3 - substantive advancement of Seamen after 6 years

Educational Attainment on Entry	Recruiting Test Scores	Proportion still Serving at each Rate			N
		Able	Leading	PO +	
With GCE	High	.50	.37	.12	50
Without GCE	High	.65	.34	-	50
"	Medium	.81	.18	-	49
"	Low	.94	.06	-	54

(from Barnecutt, 1968)

Table 4 - Days in Detention for Seamen over 6 years

Educational Attainment on Entry	Recruiting Test Scores	Average Days in Detention	N
With GCE	High	1.6	50
Without GCE	High	9.1	50
"	Medium	3.5	49
"	Low	6.1	54

(from Barnecutt, 1968)

** In terms of having one or more passes in the General Certificate of Education (or equivalent) at Ordinary level which is taken by [illegible] school children at the end of their compulsory education (ie at age 15 or 16).

From Tables 2-4 one can see that a consideration of both Recruiting Test scores and civilian educational attainments produce better predictions than test scores alone (cf Wool 1968; Plag et al 1971). Table 4 especially, points to the possibility of an important personality or attitudinal variable. It also shows that there are some criteria of great operational significance which are not necessarily covered by the steady state objectives (ie criteria concerned with quality rather than quantity).

Cost-Effectiveness

As Sands argued last year, cost-effectiveness is the single best basis on which to evaluate selection policies (cf Fisher 1969). Starting from actuarial tables of the type shown I have costed some of the alternative approaches to the selection of Artificer Apprentices (Naval Technicians) with the results shown in Tables 5 and 6.

Table 5 refers to the problem of deciding between four different systems against the criterion of the need to produce 411 Artificers at the end of training. The four systems considered are:

(1) GCE/Entry Exam (ie fill the vacancies by accepting all GCE qualified Applicants and then as many of the remainder as necessary by Entry Exam order of merit).

(2) GCE/T2 (ie as above but substituting T2 for Entry Exam).

(3) Interviewer's Mark (ie fill the vacancies by accepting as many as necessary by order of merit on the Interviewer's Mark).

(4) T2 (ie as above but substituting T2 for Interviewer's Mark).

The total number recruited is a function of the variables considered, the base rates and the selection ratios (cf Sands 1970). The approach is by successive approximations. For example:

(1) First determine the number of Applicants in the highest scoring segment (corrected for other rejection grounds if necessary, eg medical).

(2) Then refer to the actuarial tables to determine how many of these would meet the criterion (in this case complete training).

(3) If the number meeting the criterion is less than the target, accept these Applicants and move on to consider the next highest scoring group.

(4) Continue until the target is achieved (ie 411 completing training).

The outcome is an estimate for each system considered, of the quantity and quality of the Applicants needed to be recruited in order to produce 411 trained Artificers. These hypothetical Entrant cohorts were then costed for time taken to complete training and for 'benefits' accruing as a result of early completion. A 'cost per man completing' was then calculated for each, as shown in Table 5. Also shown is the number recruited in each case. Either of these could be regarded as an index of cost-effectiveness. In this case the 'cost per man completing' is preferred since it takes into account not only the base rate (ie the proportion of each cohort completing training) but also their different rates of progress (ie an additional aspect of 'risk). It is therefore more comprehensive. As Sands pointed out, however, it could be further improved by incorporating a cost factor relating to the expense of running the Selection system itself.

From Table 5 it can be seen that all four systems would achieve the target if used for the particular Applicant cohort considered but that the costs would differ. The difference between the least expensive and the most expensive is, in fact, only £9,042 ($22,000) per cohort completing (ie 411 (£3176-£3154)).

Table 5 - target 411 trained Artificers

Selection System	Cost per man Completing (£)	Number Recruited
GCE/Entry Exam	3154	461
GCE/T2	3163	465
Interviewer's Mark	3164	465
T2	3176	467

It may not be apparent that the cost-effectiveness index is sensitive to changes in the base rates (ie the validities). This can be shown by re-calculating the cost-effectiveness of these same four systems against a different criterion. As was pointed out, against the criterion of overall wastage (411 completing) there was only some £9000 ($22,000) difference in costs between the best and the worst systems. The main reason for this was that overall wastage was a relatively undifferentiating criterion for Artificers (ie almost 90% of all Recruits completed whatever

their scores, if given time). However, if a more differentiating criterion is chosen the picture changes. An example is provided by using 'rate of progress' as a criterion. In this case an inspection of the actuarial tables revealed that some systems were highly valid (ie they clearly differentiated people into bands of risk) whereas others were relatively unpredictive (ie high scorers performed no better than low scorers). Table 6 shows that the cost-effectiveness index allows such differences to be quantified.

Table 6 shows that all four systems would again achieve the target if used for the particular Applicant cohort considered but that the costs would differ. The difference between the least expensive and the most expensive is now £139,216 ($338,000) per cohort completing.

Table 6 - target 308 Artificers completing in minimum time.

Selection System	Cost per man Completing (£)	Number Recruited
GCE/T2	5265	466
T2	5334	472
GCE/Entry Exam	5423	480
Interviewer's Mark	5717	506

The fact that there are many ways of meeting planning objectives is further illustrated by Table 7. In this case I have costed two variations on the T2 based system:

1. T2 (Highest) involved filling the vacancies by T2 order of merit.
2. T2 (Lowest) involved filling the vacancies by T2 reversed order of merit (ie recruiting the 'worst' on T2).

As can be seen (Table 7) both policies would meet the target but at different cost - in this case a total of £46,854 ($114,000) per cohort completing separates the two.

Table 7 - target 411 trained Artificers

Selection System	Cost per man Completing (£)	Number Recruited
T2 (Highest)	3176	467
T2 (Lowest)	3290	456

It may be that this cost-effectiveness difference could provide a useful additional measure of validity with regard to a given situation. Certainly it would be advisable to always calculate it as a check that the common assumption 'highest is best' really pertains. On this point one should note that the actuarial tables themselves do not necessarily provide an acid test since they reflect base rates and not selection ratios. In other words, an actuarial table might indicate a difference in success rate between high and low scorers which might suggest the use of a particular selection variable whereas an examination of the selection ratios of the various segments might show that, although high scorers, in principle, do better, there are so few amongst the actual Applicants that elaborate selection is unworthwhile. The cost-effectiveness of random selection, if it can be calculated would provide yet another base line.

One other feature of the Artificer study was the finding that the Selection Board Interview was relatively valueless as a predictor for dividing people into different bands of risk with respect to training and subsequent success. Much the same finding was reported by a namesake of mine (Cdr K E Gardner 1970) in a series of studies he made in the area of Officer selection for the Royal Navy.

Gardner's analyses showed that the abilities associated with long term success were different from those associated with short term success and that these relevant abilities varied from Branch to Branch. Training success insofar as it was predictable and as measured by course results was almost wholly associated with intelligence. Successful Engineer Officers tended to have high spatial-mechanical ability whereas successful Seamen Officers tended to have high verbal-educational ability. Long term managerial success, on the other hand, as measured by promotion to Commander, required a combination of high verbal-educational ability and low spatial-mechanical ability irrespective of Branch (ie some of the qualities of importance to short term success were irrelevant or even a handicap to subsequent long term success).

A judgemental mark based on observation of group tasks and interviews, was not a significant predictor of success either in the short or long term. However, the possibility remains that this mark might serve to ensure that only those candidates who can meet the demands of life in Dartmouth - the Britannia Royal Naval College - will be sent there. Thus, it does not necessarily follow that a selection system based entirely on test scores and educational attainments would be successful - it might very easily lead to considerable upheaval and chaos in the new-entry establishments for Officers add Ratings alike. Intuitively one feels that this is likely. Unfortunately, neither Gardner KE, nor myself have been able to put it to the test.

One final comment about Gardner's study, it is also noteworthy that he showed the potential usefulness of discriminant analysis for selection research. It would be interesting to know of studies done explicitly to evaluate the relative strengths and weaknesses of the various techniques for actuarial prediction.

'Standard Man'

Tables 5-7 showed that selection policies can differ in respect of the number of men they need to recruit in order to meet their output criteria. This being so, what is the relevance of specifying recruiting targets in input terms?

The short answer would seem to be that recruiting targets of the traditional type (ie input quotas) are justified only in 'recurrent situations' where the type of population, the conditions of testing and the nature of the criterion are all stable. Where this is not the case it may be more profitable to specify targets in output terms (cf Tables 5-7). To this end I tentatively offer the concept of a 'standard' man.

For selection purposes a Recruit's 'standard man' index would be defined as his estimated chance of achieving a given criterion. Of course, there are problems in this, not least being the question of what criterion should be the reference - although presumably, any given set of Applicant characteristics could be associated with a variety of criteria. In other words, conceivably a man could be classified as standard .7, say, for one criterion but standard .1, say, for another (eg a recruit potentially good at academic subjects but potentially no good at firing a rifle). Nevertheless, in the case of a single criterion, at least, it would be possible to specify the target in terms of 'standard men'. It would then be a simple matter to compute the likely yield at a criterion

for a given Applicant or Recruit population or alternatively, to compute the preferred mix of Recruits in order to achieve a given criterion in the face of a given Applicant population. It might even be possible to build in a cost factor - certainly some of the basic data exists (cf Plag et al 1971).

Incidentally, I would like to point out an interesting complication, namely, that simple head counts of Entrants would cease to be a useful measure of success in recruiting. For instance, 100 men with 'standard man' value of 1.0 would be, from the point of view of the target, the same as 200 men with a value of .5. In other words these two recruitments would be unequal in size but equal for planning purposes - a statement which, I imagine, will seem very paradoxical to some. Note, however, that the costs incurred by these two groups would be different. It might be desirable, therefore, to build a cost factor into the 'standard man' index so as to avoid unnecessary confusion.

System Monitoring

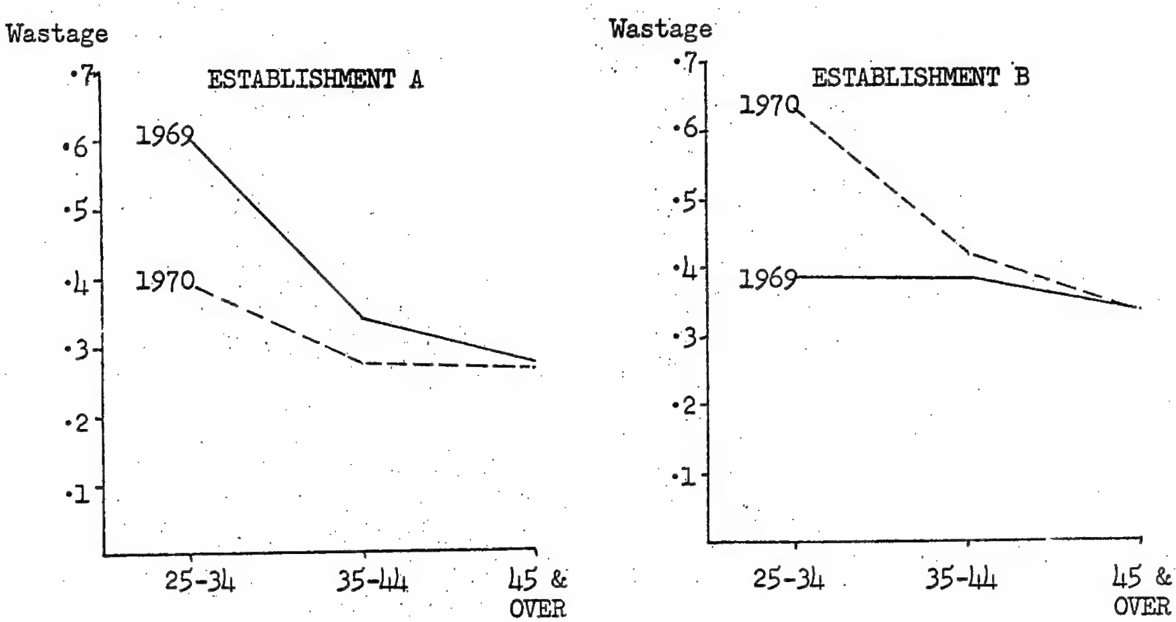
A lot has been said already about the interaction of changing policies or populations or conditions. Clearly, any actuarial approach to selection will fail if it is not kept up-to-date for, as Vernon (1964) pointed out, given their reliance on 'recurrent situations', actuarial approaches may, in practice, turn out to be too inflexible. Hopefully this could be avoided by regular monitoring and up-dating. Figures 2-4 illustrate the need for this in terms of wastage from two entry establishments (ie 'Total Wastage' and its components 'Voluntary Withdrawal' and 'Discharged Unsuitable etc'*) amongst Seamen recruits without GCE 'O' levels or equivalent (cf Tables 2-4).

---

* The majority of wastage other than 'voluntary withdrawal' was a result of being discharged for failing some part of training. Of course, the extent to which 'voluntary' actions are entirely free from organisational persuasion is always open to question.

Figure 2 shows that the relationships between tests and total wastage have not remained constant and, further, that the two establishments have moved in different directions.

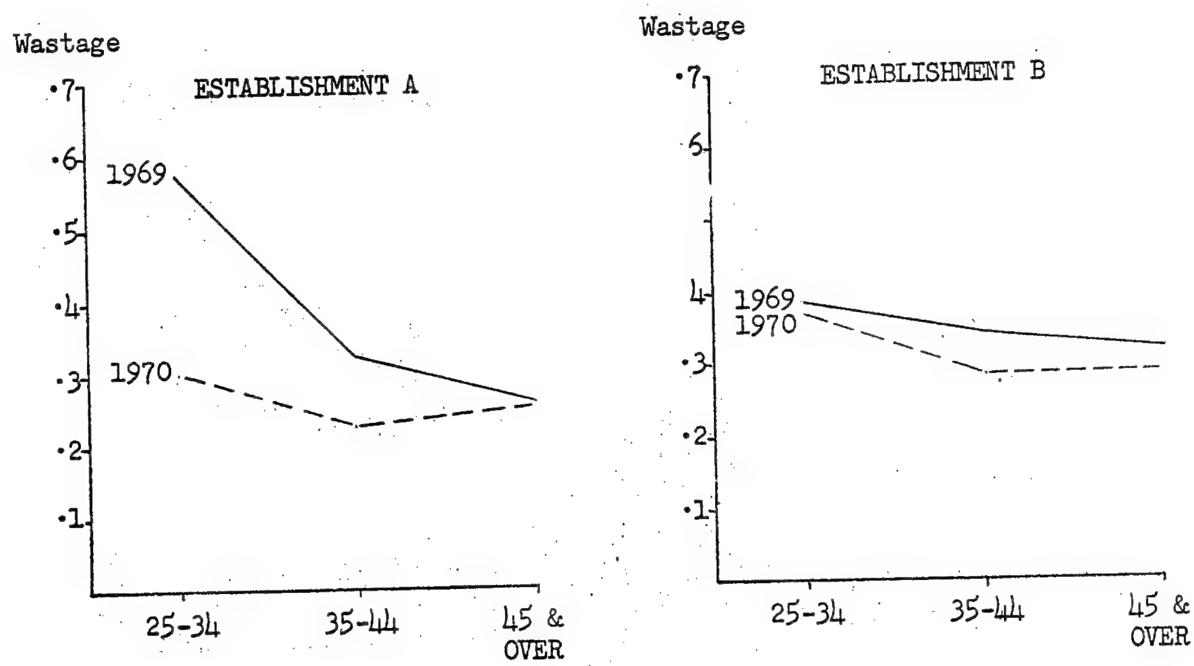
Figure 2 - Total Wastage (Seamen)



(Source: Gardner, 1971)

Figure 3 shows that voluntary withdrawals, especially amongst the lowest scorers, has declined in Establishment A but not - to all practical purposes - in B.

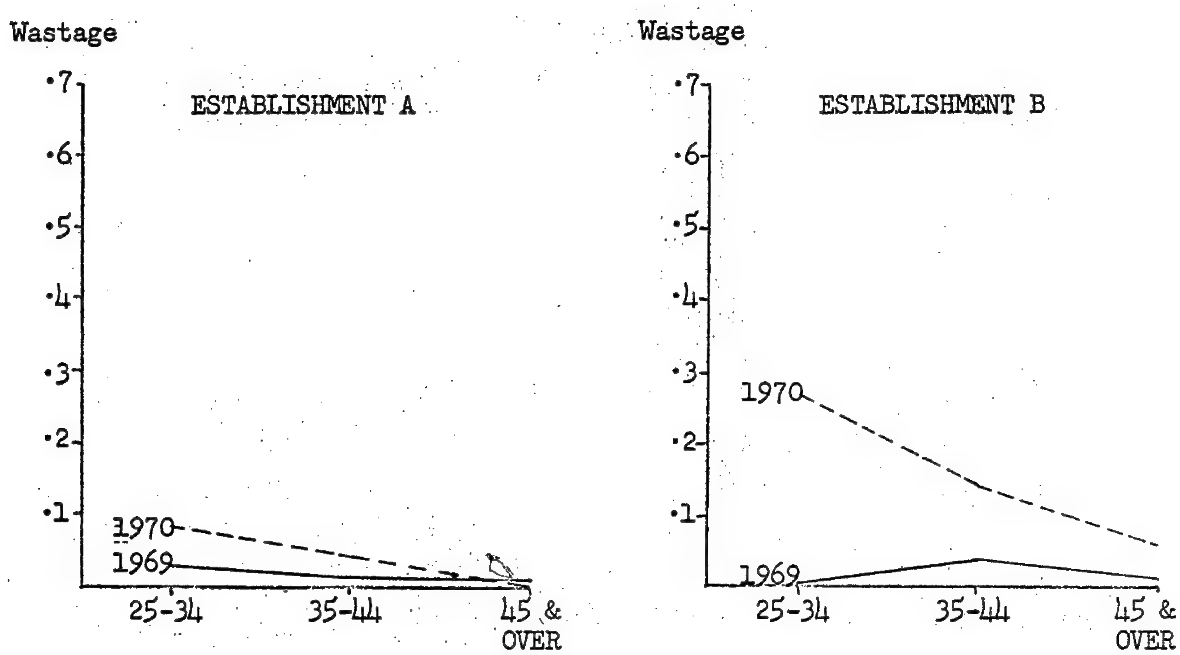
Figure 3 - Voluntary Withdrawal (Seamen)



(Source: Gardner, 1971)

Conversely (Figure 4) wastage for reasons other than Voluntary withdrawal amongst lower scorers has risen only marginally in Establishment A but almost trebled in B.

Figure 4 - Discharged 'Unsuitable' etc (Seamen)



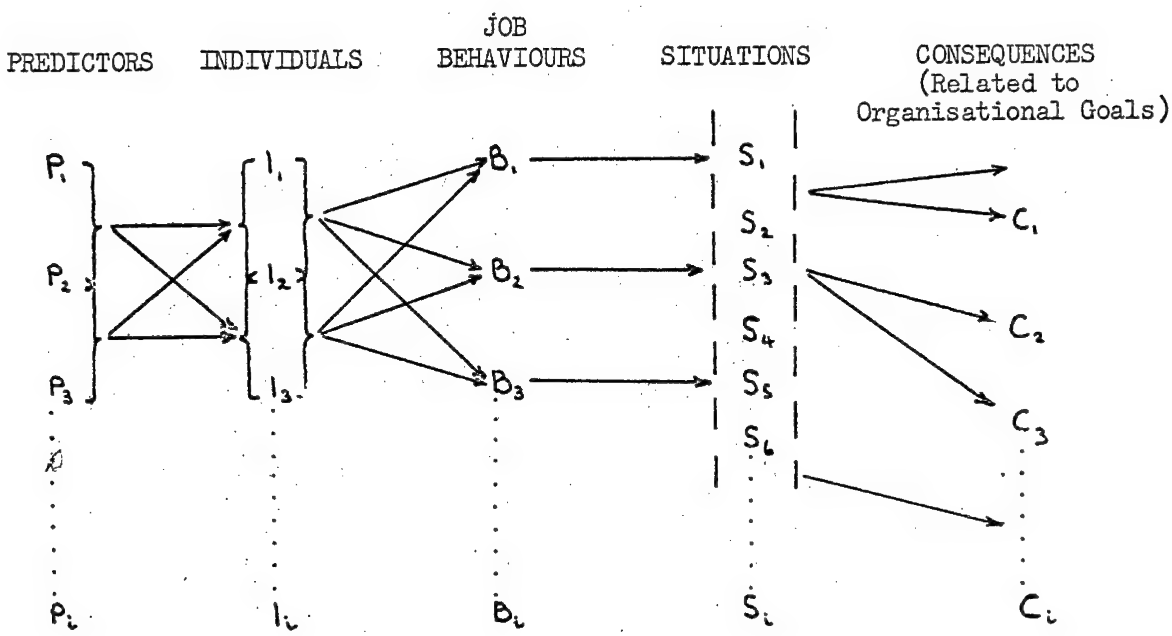
(Source: Gardner, 1971)

Changes such as these would negate any planning unless they could be predicted on the basis of other factors. What these might be is a matter for research but some possibilities to consider are: regular attitude surveys to determine trends; or supplementary tests after Entry, allied to progress reports. There is, of course, also the need to be aware of policy

decisions made elsewhere than the selection stage which might affect conditions of service, etc., in the particular establishments. In other words, I am advocating in 1971 what Dunnette advocated in 1963! Dunnette, it will be remembered, presented a model for test validation and selection research (Figure 5) in which he drew attention to the fact that there were important interactions between predictors, individuals, job behaviours, situations and organisational consequences.

> "... the modified prediction model taken account of the complex interactions which may occur between predictors and various predictor combinations, different groups (or types) of individuals, different behaviours on the job, and the consequences of these behaviours relative to the goals of the organisation. The model permits the possibility of predictors being differently useful for predicting the behaviours of different subsets of individuals. Further, it shows that similar job behaviours may be predictable by quite different patterns of interaction between groupings of predictors and individuals or even that the same level of performance on predictors can lead to substantially different patterns of job behaviour for different individuals. Finally the model recognizes the annoying reality that the same or similar job behaviours can, after passing through the situational filter, lead to quite different organizational consequences."

Figure 5 - Dunnette's Model for Test Validation and Selection Research



(Source: Dunnette, 1963)

One can see that, in theory at least, the actuarial approach could be extended to take account of all these factors (ie tables would show the expectation of success for a given set of predictor scores given a particular population of individuals with regard to a given criterion in a given situation) although whether it could always be done in practice is another matter. Nevertheless, the approach would have the merit of automatically taking into account moderator and suppressor variables (cf Blum and Naylor 1968) provided, of course, that these were included in the initial specification. In fact, a logical consequence would be to abandon the search for a measure of overall validity (eg a correlation) since each cell of the actuarial table would be qualitatively unique (ie it would specify uniquely a given combination of predictors, populations, behaviours and situations). In this sense it would have much in common with the concept of 'conditional validities' advanced by Blum and Naylor (1968) in

their discussion of moderating variables. Thus, one would refer to the conditional validity of each cell to predict the organisational consequences in specific situations but refer to the cost-effectiveness wherever an overall judgement was required (ie the difference between a statement of validity and an overall evaluation). One should bear in mind however the inherent difficulties not least the need for very large numbers and detailed personnel records and that not all measures of job success will be reduceable to cost terms.

Finally, I want to draw attention to the fact that the discussion, so far, has concentrated on ways of maintaining the status quo - the steady state-by selection. The real challenge, however, is a wider one of identifying methods - whatever they may be - for changing manpower structures in an orderly fashion. It will be obvious, I hope, that selection is but one of the many possible approaches (cf Dunnette).

It follows that not only is there an onus on planners to take account of the inter-related variables (eg selection, training, motivation, working conditions) but equally for all the other specialists to do the same. For instance, one ought to see trainers and equipment designers, say, quoting AFQT-type scores and controlling for moderating variables in their experiments as a matter of course. How many do so? My own inclination, clearly, is to evaluate all these tools of management on the same basis. That is to say, to calculate the success rates of given men with given characteristics in given situations against given organisationally important criteria. The effectiveness of the new-style training, or whatever, could then be assessed by direct comparison with the relevant base-line(s) and, if possible, costed for its implications. In this respect, incidentally, I am encouraged by the increasing trend towards objective training (ie training based upon clear behavioural objectives). By this means training success should be more closely related to on-the-job success and, by virtue of the thereby simplified criterion problem, so too should selection.

This brings me back to my starting point (Figure 1). Currently manpower planning tends to be concerned with quantity rather than quality and tends to ignore the empirical fact that people do differ with respect to their potential for achieving organisational goals. I have suggested that this is an oversight; that an actuarial approach should lead to better planning and control of resources; and that to this end a 'good' system is a 'cost-effective' system - although it is worth remembering that there are some system features which,

even though they cannot be costed, are nevertheless desirable (eg the provision of a facility for system monitoring). Unfortunately one should also remember that these ideas have not - to my knowledge - been put to a practical test.

ACKNOWLEDGEMENTS

I am grateful for the help and advice given by Mr E Elliott and Mr A C Godwin in response to various drafts of this paper.

REFERENCES

Barnecutt P S	(1968)	"Progress of Seamen entering with Educational Certificates". SP(N)7 Memorandum 2/68.
Blum M L and Naylor J C	(1968)	"Industrial Psychology" Harper and Row New York, 1968.
Dunnette M D	(1963)	"A Modified Model for Test Validation and Selection Research", Journal of Applied Psychology, Vol 47, 1963, pp 317-23.
Fisher F M	(1969)	"Aspects of Cost-Benefit Analysis in Defence Manpower Planning" see Wilson N A B (ed)
Gardner A	(1971)	"RT and Wastage Rates" SP(N) Memorandum 2/71
Gardner (RN) Cdr K E	(1970)	"A Long-Term Follow-up of Naval Officers' Careers using Multivariate Techniques" unpublished M Phil thesis, City University, London.
Jones E	(1969)	"Application of Actuarial Techniques to Officer Career Planning" see Wilson N A B (ed)
Plag J A, Goffman J M and Phelan J D	(1971)	"Predicting the Effectiveness of New Mental Standards Enlistees in the US Marine Corps" Navy Medical Neuropsychiatric Research Unit, San Diego, California 92152. Report No 71-42.
Sands W A	(1970)	"Cost of Attaining Personnel Requirements (CAPER) Model" in Proceedings of the 12th Annual Conference of the Military Testing Association, Enlisted Evaluation Center, Fort Benjamin Harrison, Indiana 46249.
SP Report QC 33	(1957)	"The Relation between Entry Tests and Advancement in the Seaman Branch".
Vernon P E	(1964)	"Personality Assessment" Methuen, London John Wiley, New York 1964.
Wilson N A B (ed)	(1969)	"Manpower Research", The English Universities Press, London, 1969.
Wool H	(1968)	"The Military Specialist: skilled manpower for the Armed Forces". Johns Hopkins Press, Baltimore 1968.

Milton H. Maier
U. S. Army BESRL
16 September 1971

An Improved Army Classification Battery

Each of the armed services is faced with the problem of selecting, classifying, and assigning to training and jobs large number of young men who enter the service. Most of the men have limited work experience and little technical training; thus little information is readily available about the jobs they can best fill while on active duty. Selection and classification tests have been developed and used over the years to measure the potential of the young men to perform in the large variety of jobs open to recruits. In the Army the number is several hundred jobs. The tests are an efficient and accurate means of assessing the potential to succeed in each of the jobs or associated training courses, and permit an effective match between the needs of the service and the capabilities of the new recruits.

The Army Classification Battery (ACB) has been used for over 20 years in assigning men to their job training courses. As the Army moves from an induction input to an all volunteer input, the ACB will be used increasingly to help make selection decisions as well. In the various enlistment options available, mental standards are set in terms of the Armed Forces Qualification Test (AFQT) and aptitude area scores, derived from the ACB or its counterpart,the Army Qualification Battery (AQB). Using the test battery for selection imposes somewhat different requirements than using it only for classification. If the tests were used only for classification and assignment, then the differences between the aptitude scores would be the critical factor. We would want to know if the man

had more potential as, say, a mechanic or clerk; but because he already has been selected into the service on the basis of another test, such as the AFQT, we would assume that he is at least minimally qualified in both areas. If the test battery is used for both selection and classification, then we want to know not only the differences in potential, but also whether he is mentally qualified in the different areas.

The distinction between selection and classification was brought home forcibly to the Army when the mental standards were lowered for Project 100,000 in 1966. Up until that time all the input had the minimal levels of literacy and general mental ability because the mental standards were high. The aptitude area scores obtained from the ACB were used to reveal the job areas in which the men could be best utilized. Men could be assigned to the different areas on the basis of their aptitude area scores with reasonable confidence that they would perform satisfactorily.

When selection standards were lowered to accept men with lower mental ability, the level of general ability was changed sufficiently to create problems. Some Army schools were receiving too many men who could not absorb the highly technical material because they were too low on general ability. We developed an interim solution to make the aptitude area scores more suitable with the new levels of input, but it was not implemented operationally because of the manpower implications. We were developing a new classification battery at the time, which is expected to be ready for operational use late in 1972. The aptitude area scores obtained from the new ACB are more suitable for the dual purpose of selection and classification.

Development of the New ACB

One need for a new system to select and classify Army enlisted men became apparent with the lowering of mental standards. Another need for a modification of the test battery arose from the passage of time and the associated changes in our culture and the Army training courses. Some of the test items date back to the 1950's, and some of these seem out of place today. For example, the German general, Rommel, The Desert Fox, no longer receives much popular attention, and television picture tubes have changed design as compared to 20 years ago. In addition, the Army technology has changed with more sophisticated equipment. The training requirements have been increased to enable the men to cope with the more technical machines and concepts.

Starting over a decade ago, BESRL began an extensive research program to develop a new version of the ACB. The research objectives were to update the tests that were becoming obsolescent, to find content more suitable to the newer types of job training courses and to find new kinds of tests that would increase the predictive accuracy of the battery.

The research program was successful in that a new version was developed which does meet more effectively the needs of the modern Army. The relationship between the tests in the old and new ACB is shown in Table 1.

The New Army Classification Battery

General Ability Tests. The new ACB has five tests of general ability, three common to the new and previous battery, and two added tests, Mathematics Knowledge and Science Knowledge. The Word Knowledge and Arithmetic Reasoning tests are changed from the original only in having been shortened to provide more efficient measurement. The

TABLE 1

CONTENT OF NEW AND PRIOR ARMY CLASSIFICATION BATTERIES

NEW ACB		OLD ACB		CHANGE
GENERAL ABILITY TESTS				
ARITHMETIC REASONING	(AR)	ARITHMETIC REASONING	(AR)	SHORTENED
WORD KNOWLEDGE	(WK)	VERBAL	(VE)	SHORTENED
GENERAL INFORMATION	(GI)	GENERAL INFORMATION	(GIT)	UPDATED AND SHORTENED
MATHEMATICS KNOWLEDGE	(MK)			ADDED
SCIENCE KNOWLEDGE	(SK)			ADDED
MECHANICAL ABILITY TESTS				
ELECTRONICS INFORMATION	(EI)	ELECTRONICS INFORMATION	(ELI)	UPDATED
MECHANICAL COMPREHENSION	(MC)	MECHANICAL APTITUDE	(MA)	UPDATED
AUTOMOTIVE INFORMATION	(AI)	AUTOMOTIVE INFORMATION	(AI)	SHORTENED
TRADE INFORMATION	(TI)			ADDED
		SHOP MECHANICS	(SM)	DROPPED
PERCEPTUAL ABILITY TESTS				
PATTERN ANALYSIS	(PA)	PATTERN ANALYSIS	(PA)	UPDATED
AUDITORY PERCEPTION	(AP)	ARMY RADIO CODE APTITUDE	(ARC)	NONE
		ARMY CLERICAL SPEED	(ACS)	DROPPED
ATTENTION TO DETAIL	(AD)			ADDED
INTEREST TESTS				
COMBAT SCALE	(CC)	CLASSIFICATION INVENTORY	(CI)	UPDATED AND SHORTENED
ATTENTIVENESS SCALE	(CA)			ADDED
ELECTRONICS SCALE	(CE)			ADDED
MAINTENANCE SCALE	(CM)			

General Information test, updated and shortened, has shifted from its function as primarily a combat selector to serve as a measure of the general ability required of good performers in selected noncombat MOS as well as in artillery. The Mathematics Knowledge and Science Knowledge tests were added to expand coverage in this important aptitude domain. Each of the five tests measures a different aspect of general ability. The Word Knowledge, Arithmetic Reasoning, and General Information tests cover skills and knowledge that can be acquired in or out of school. The other two tests cover abilities taught in formal school courses. All five tests measure aptitudes required in a wide variety of jobs and situations.

Mechanical Ability. Four mechanical ability tests are included in both batteries. The Automotive Information Test was shortened for the new battery. The Shop Mechanics Test was dropped and replaced by Trade Information. Content of the Electronics Information Test was updated. The Mechanical Aptitude Test was updated and the title changed to Mechanical Comprehension. The new tests have the advantages that the content is up to date, the tests are more valid, and all are shorter.

Perceptual ability. The three tests of perceptual ability require no reading or writing skills but do require ability to perceive certain kinds of stimuli--geometrical patterns, and auditory and visual symbols. The new version of the Pattern Analysis Test, which requires visualization of three-dimensional form, is shorter than the previous one. The Army Radio Code Aptitude Test has a new title, Auditory Perception, but otherwise remains the same. The more inclusive title reflects the finding that the test is useful for jobs other than radio operator--jobs that require the ability to listen attentively. The Army Clerical Speed

test was replaced by Attention to Detail, which is more widely useful and easier to administer.

Self-Description Test. An expanded version of the Classification Inventory, long used to identify men who will make good combat soldiers, was introduced. Four separate measures are obtained from this test: Scale CC corresponds to the previous Classification Inventory score used to identify combat infantrymen, but it has been updated and shortened. Scale CA is a measure of attentiveness, a useful predictor for a variety of jobs--clerical, artillery, missile crewman, for example. Scale CE (electronics) and Scale CM (maintenance) are related to specific job families; both help identify repairmen who will be successful in the relevant area.

Grouping the MOS

The development of new tests does not by itself result in a new classification system. A critical component is that of grouping the jobs into relatively homogeneous clusters or families. As already mentioned, there are several hundred jobs potentially open to the Army recruit, and some of these jobs are more alike than others. The Army philosophy, with some support from research data, is that the jobs should be grouped into a manageable number of categories, which is in the range of about 8 to 12. The jobs within a group should require similar skills and aptitudes for success, and be as different as possible from the jobs in other families. Thus the electronics repair jobs form one cluster, which is different from the clerical-administrative jobs.

The grouping of the jobs was accomplished by computing the validity of each test for each sample separately. Since we had over 30 tests and 100 samples, there were well over 3000 validity coefficients.

We examined the validity profiles for each sample, and grouped those jobs together that tended to require the same aptitudes and interests for success. We had some help in this grouping process from the structure of Army jobs already in existence. We found that with some exceptions we could base our grouping on that used operationally.

The final result of grouping the jobs is shown in Table 2. We ended up with nine MOS groups, shown in the left column; some representative jobs in each group are shown in the right column. Each of these groups was developed on an empirical basis. The MOS were grouped together only if our data showed that they were similar in terms of the interests and aptitudes required for success and different from the MOS in other groups.

The first MOS group, called CO for Combat, includes the infantryman, armor crewman, and combat engineer. The second group, FA for Field Artillery, includes the field cannon and rocket artillery jobs. The third group, EL for Electronics Repairs, includes all electronics and electrical repairmen. An attempt was made to keep the electronic and electrical maintenance MOS separate, but our data did not support such a distinction, and we combined them.

The fourth group, OF for Operators-Food, includes a seemingly diverse collection of jobs: missile crewmen, cooks, and drivers. The grouping emerged from our data, and their common feature seems to be a requirement for low level maintenance skills. The missile crewman has to handle equipment, the driver has to check out vehicles, and the cook has field stoves to set up. The next groupSC, for surveillance and communications, includes radio operators, communication center specialists, and switchboard operators. The MOS involve receiving and processing information;

TABLE 2

REPRESENTATIVE JOBS IN NEW MOS GROUPS

MOS GROUP		REPRESENTATIVE JOBS
COMBAT	(CO)	INFANTRY, ARMOR, COMBAT ENGINEER
FIELD ARTILLERY	(FA)	FIELD CANNON AND ROCKET ARTILLERY
ELECTRONIC REPAIR	(EL)	MISSILES AND AIR DEFENSE REPAIRMEN, TACTICAL ELECTRONIC AND FIXED PLANT COMMUNICATIONS REPAIRMEN
OPERATORS/FOOD	(OF)	MISSILES AND AIR DEFENSE CREWMEN, DRIVER, FOOD SERVICES
SURVEILLANCE/COMMUNICATIONS	(SC)	TARGET ACQUISITION AND COMBAT SURVEILLANCE AND COMMUNICATION OPERATIONS
MECHANICAL MAINTENANCE	(MM)	MOTOR AND AIRCRAFT MAINTENANCE, RAILWAYS
GENERAL MAINTENANCE	(GM)	CONSTRUCTION, UTILITIES, CHEMICAL, MARINE, PETROLEUM
CLERICAL	(CL)	ADMINISTRATIVE, FINANCE, SUPPLY
SKILLED TECHNICAL	(ST)	MEDICAL, MILITARY POLICE, DATA PROCESSING, AIR CONTROL, TOPOGRAPHY AND PRINTING, INFORMATION AND AUDIO VISUAL

the common element seems to be a requirement for perceptual ability, both auditory and spatial.

There are two maintenance groups, MM for Mechanical Maintenance and GM for general maintenance. The MM group includes motor mechanics, aircraft maintenance and railway jobs. GM covers a variety of jobs, such as construction, utilities, marine, chemical, and petroleum.

The final two groups, CL for clerical and ST for skilled technical, are familiar to Army personnel. CL includes the administrative, finance, and supply jobs. ST is similar to the old GT area, and includes medics, military policemen, and intelligence specialists.

The New Aptitude Area Composites

The final step in developing a new classification system is to find the weights to assign to each test for each MOS group. The weights were obtained by determining which tests contributed most to predicting success in each area. In selecting the tests, we first selected the test that was most valid; then we added the test that made the second largest contribution to validity. We continued the selection process until the remaining tests made little contribution to increasing the accuracy of prediction. Generally we had to select 4 or 5 tests for each MOS before we exhausted the validity of the battery; in two of the areas we needed only three tests. The tests selected for each MOS area were assigned a weight of one; we found through extensive simulation studies on the computer that simple unit weights were as effective as more elaborate weighting schemes. Those tests not selected were assigned a weight of zero. The tests used for the MOS groups are shown in Table 3.

Table 3

NEW APTITUDE AREA COMPOSITES

Test		CO	FA	EL	OF	SC	MM	GM	CL	ST	GT
		Aptitude Area Composites									
General Ability Tests											
Arithmetic Reasoning	(AR)	AR	AR	AR		AR		AR	AR	AR	AR
General Information	(GI)		GI		GI						
Mathematics Knowledge	(MK)		MK				MK			MK	
Word Knowledge	(WK)					WK			WK		WK
Science Knowledge	(SK)							SK		SK	
Mechanical Ability Tests											
Trade Information	(TI)	TI		TI			TI				
Electronics Information	(EI)		EI	EI			EI				
Mechanical Comprehension	(MC)			MC		MC		MC			
Automotive Information	(AI)				AI		AI	AI			
Perceptual Ability											
Pattern Analysis	(PA)	PA				PA					
Attention to Detail	(AD)	AD							AD		
Auditory Perception	(AP)					AP					
Self Description											
Combat Scale	(CC)	CC									
Attentiveness Scale	(CA)		CA		CA				CA		
Electronics Scale	(CE)			CE							
Maintenance Scale	(CM)						CM				

Our research showed that the good combat soldier needs general ability, measured by the Arithmetic Reasoning Test, mechanical ability, measured by the Trade Information Test, to handle his weapons and equipment; perceptual ability measured by the Pattern Analysis and Attention to Detail tests, to orient himself in the terrain and observe his environment, and finally, an interest in outdoor masculine activities, coupled with self-confidence, measured by the Combat scale (CC) of the Classification Inventory.

The artilleryman, in comparison was found to require more mathematical ability. Therefore, scores from both the Arithmetic Reasoning Test and the Mathematics Knowledge Test enter into the Field Artillery (FA) Aptitude Area. A further measure of general ability is contributed by the General Information Test. Mechanical ability, measured by the Electronics Information test, and an interest in details, measured by the Attentiveness (CA) scale of the Classification Inventory, complete the picture for the artilleryman.

A similar analysis can be made for each MOS group. In all cases the tests that were selected for the composites made sense based on what we know about the jobs in each group.

A word should be said about the final composite in the list, the familiar General Technical (GT) Aptitude Area, composed of the Arithmetic Reasoning and Word Knowledge (Verbal) tests. In the old system, the GT score is used both to select men for general technical MOS and to determine which men are eligible to take additional tests such as the Officer Candidate Test. The function of selector for MOS group is shifted to the ST composite. The function of determining eligibility for additional testing continues to be filled by the combination of Arithmetic Reasoning and Word Knowledge Tests. The label GT is retained.

Evaluation of the New Classification System

The new Army classification system was carefully evaluated to estimate how much improvement would be realized over the old system. The conclusions we have reached are that academic attrition in job training courses would be reduced by about 20 percent; that the number of marginal performers, that is, men who barely pass the training course, would be reduced by 20 percent, and that the number of superior performers would increase by 15 percent. The procedures were to run extensive simulation studies on our computer; the details will be given in a forthcoming BESRL technical report. In obtaining these estimates of improvement over the old classification system, the quality of men coming into the Army was assumed to be exactly the same for both classification systems. The improvement in performance can be realized because the new system does a better job of getting the right man into that area where he can perform best.

The new ACB would result in even greater benefit when the improved selection is considered. Applicants for specific jobs need to qualify on several aptitude scores. Since the new aptitude scores are more accurate measures of potential than the old scores, there is greater assurance than men who meet these requirements can be trained to the point of acceptable competence in an Army job.

In summary, a new Army Classification Battery has been developed that will result in improved selection and classification of enlisted men, and it will be ready for operational use late in 1972. New measures of interest and general mental ability have been added to the battery, and new combinations of tests have been developed to measure more accurately the potential to perform in training and on the job. The effect of the new system is to screen out more of the men who would be likely failures and to utilize more effectively the talents and interests of the men who are accepted for Army service.

Contemporary Validation Approaches and a Discussion of Seasonal, Regional and Language Differences in the CAF Classification Battery - Men

Harvey A. Skinner[1]

Canadian Forces Personnel Applied Research Unit

A general program is currently in progress at CFPARU to assess and enhance utility of the Classification Battery - Men (CBM) and to examine the feasibility of a more inclusive selection - classification model. One component is the empirical evaluation of a more individualized selection strategy through a moderator variable approach.

Following a brief discussion of recent trends in test validation and selection research, this paper presents a summary of the basic issues of moderator variables and their potential application in applied settings. In addition, the experimental strategy of this project is outlined and available results from Phase I are examined with respect to seasonal, regional and language differences in the 1971 enrolment population.

Recent Trends

Limitations of the classic validation model have stimulated newer, more comprehensive strategies which recognize the complexities involved in predicting human behaviour as well as the necessity of examining the problem of prediction in a more embracing, system context. One frequently referenced model (Dunnette, 1963, 1966) embodies the complex interactions among 1) predictor subsets, 2) different types of individuals, 3) on-job behaviour patterns, 4) situational differences and 5) organizational consequences and goals. Through empirical investigations, optimal groupings are sought which maximize overall efficiency and

[1] The author wishes to acknowledge Glenn M. Rampton and Wayne E. Keates for their helpful suggestions during the progress of this research. In addition, a special thank you is extended to Mr J.A. Doran and his staff for their faithful assistance in documenting and processing the data.

foster a deeper understanding of selection system components and their interrelationships. Likewise, Rundquist (1969) purports the need for fewer statistics and more throught and experimentation. Personnel selection, he argues, must be viewed in as large a context as possible since man is an open system working as part of a larger system. Emphasis has consequently shifted from simply developing validity coefficients to a consideration of utility, a universal concept which incorporates not only measurement accuracy but also the relative importance of personnel decisions, manpower quotas, expenses associated with operating the selection system and costs accruing from selection errors.

Moderator Variables

One contemporary strategy which has attracted considerable interest is the division of the heterogeneous candidate population into more homogeneous subgroups through the identification and employment of moderator variables. These variables operate indirectly by interacting with the relationship between predictor and criteria variables such that a specific test or test battery may have differential validity for certain subsets of individuals. Compared with the total population, each subgroup is viewed as being more homogeneous in relevant psychological variables.

Some semantic ambiguity has developed since Saunders (1955) originally defined the term moderator variable in the moderated regression sense because this rubric has been applied to various other methodological approaches. However, Zedeck's (1971b) recent synthesis helps clarify the equivocation by classifying the varied concepts "in terms of whether they lead to differential validity, or lead to differential predictability, or involve moderated regression techniques."

First, differential validity techniques employ either qualitative or quantitative variables to factor the total sample population into subgroups where the validity coefficient for one subgroup is significantly different from

the validity coefficient for another subgroup. Thus, subsets are disclosed for which certain predictors are more (or less) appropriate. Frederiksen and Melville (1954), for example, used two measures of compulsivity to identify compulsive and less compulsive male engineering students. For the less compulsive group they reported significantly higher correlations between 5 of 10 scales on the Strong Vocational Interest Blank and grade-point averages.

A second approach, initiated by Ghiselli (1956) attempts to empirically elucidate subgroups who differ in predictability. That is, the absolute difference between the standardized predictor and criterion scores, $|D|$, serves as a predictability index. A high value of $|D|$ identifies individuals who deviate from the regression line and for whom a specific predictor instrument is inappropriate. Ghiselli (1963) proposes that moderators distribute individuals along a continuum - "at the one extreme, then, error of prediction is smaller and test validity higher and at the other, error is larger and test validity lower." A subgroup is comprised of those individuals who fall near the same point on the continuum. This is a departure from classic psychometric theory which assumes that the standard error of measurement is the same for all individuals. Extensions of Ghiselli's technique include quadrant analysis (Hobart and Dunnette, 1967) and multi-predictable group validation (Zedeck, 1971a).

The third approach originates from Saunders' (1955) paper where one moderated regression equation, developed for the total group, improves prediction over that of ordinary multiple regression. This moderated equation contains a higher order cross product term between moderator and predictor variables such that the beta weights instead of being constant are linear functions of the moderator.

Limitations

Although the moderator variable approach has demonstrated some success in the literature, one major limitation is the specificity of the findings. For

example, Stricker (1966) in a replication and extension of two earlier studies (Frederiksen and Melville, 1954; Frederiksen and Gilbert, 1960) also found that compulsivity moderated the relationship between Strong interest scales and grade-point average for male engineering students. But, this moderating effect did not generalize to male and female liberal arts and science students. A second criticism involves the practical constraints (especially in applied research settings) of employing the tedious 'one variable at a time' approach in investigating potential moderators.

In response to these limitations, recent research efforts have focused upon 1) examining multiple variables via multivariate procedures and 2) developing computer programs to quickly and systematically identify viable moderators. One computer program developed by Rock, Barone and Linn (1967) was successfully applied in the prediction of law school success by Klein, Rock and Evans (1968). Employing five potential moderators simultaneously (age, previous field of study, preparation for law school, father's occupation, and time when law school was first seriously considered) they found four homogeneous subgroups exhibiting differential validity. In a similar study, Zedeck, Cranny, Vale and Smith (1971) coined the term 'joint' moderators which they define "as two or more variables, quantitative or qualitative, that interact to influence a validity coefficient." However, they were unsuccessful at identifying 'joint' moderators (anxiety and study habits) in their study, although anxiety singly moderated the relationship between grade-point average and American College Test scores and study habits functioned as an independent predictor.

Certainly, research emphasis must be shifted from simply identifying new moderators towards gaining an understanding of the fundamental operation of moderator variables. Only when general principles are empirically established and efficient, systematic techniques for testing potential moderators are perfected, will the full potential of this approach be realized in applied

research.

'Fair' Testing

One practical advantage of moderator variable research is that this approach enables the detection of selection bias for or against specific subgroups which would otherwise remain masked when considering the population as a whole.

Recently, in the social sciences, attention has focused upon environmental factors that influence or moderate behaviour as reflected in test performance. Investigators have examined the effects of cultural disadvantage, ethnic group membership, education, socio-economic level, etc. This research emphasis, according to Wolf (1966), has been long overdue since:

> All theories of learning and behaviour make provision for the influence of the environment on the development of human characteristics, but --- we have not had a corresponding emphasis in our measurement procedures.

Parallel to this scientific interest, there has developed a surging public demand for equal opportunities and 'fair' testing within the educational and personnel selection fields. Evidence of this concern is provided by passing of the American Civil Rights Act of 1964 (in particular the Tower amendment) and the frequently referenced Motorola case (see Ash, 1966).

The following discussion focuses solely upon selection bias and assumes that the criteria measures are both accurate and relevant. Consideration is excluded of any systematic bias which may be operating throughout the selection-training-job performance spectrum. As Guion (1967) points out: "discrimination has been a part of the culture so long that even the well-intentioned may discriminate inadvertently." This is not the millenium; we cannot eliminate all bias. However, within the personnel selection domain, positive steps can be taken to elucidate selection bias and to devise means of erasing or at least minimizing demonstrated inequities.

Since test scores only indicate an individual's performance level at that given point in time, a candidate with (for example) more experience in taking paper and pencil tests - test wiseness - may score higher than another candidate from a different environment with less experience in testing situations, even though the two candidates may be similar in innate intelligence, aptitudes, and abilities. This test difference is valid as a predictor of future performance only if the higher scoring candidate consistently rates superior in training, job performance, or any other criteria one is trying to predict. Fair discrimination among candidates, where candidates' (unbiased) criteria scores are accurately predicted by their predictor measures, is desired. Unfair discrimination exists, according to Guion (1966), when candidates "with equal probabilities of success on the job have unequal probabilities of being hired for the job." In other words, a selection battery is 1) biased against a particular subgroup if predictor scores continually underpredict their actual criterion measures and 2) biased for a subgroup if predictor scores consistently overpredict their criterion measures (see Figure 1). Although the usual connotation of bias is in the former, 'biased against' sense, the 'biased in favour of' form can be equally as important to the organization as well as to the

---

Insert Figure 1 about here

---

individual. For example, consider the consequences (in organizational expense and individual safety and happiness) of spuriously assigning a candidate to pilot training because the selection battery overpredicted his actual potential at flying.

In the literature there is surprisingly little empirical research documented which probes the existence of unfair discrimination and evaluates the effectiveness of potential procedures for adjusting these inequities. To help

fill this void, several large scale studies have recently been conducted. Kirkpatrick, Ewen, Barrett and Katzell (1968) examined six different job situations with White, Negro and Spanish ethnic groups as moderator variables. They report instances where some tests significantly correlate with certain job performance criteria in one ethnic group but not in the other and examples of differential validity between ethnic groups. Additionally, moderated regression employing ethnic group as the moderator significantly improved prediction in two studies. In a similar investigation, O'Leary, Farr and Bartlett (1970) found, of a total of 765 comparisons, 219 instances of unfairness and 281 cases where the test was valid for one group only. Both studies indicate the potential of moderator variable techniques and endorse the necessity of validating test and test batteries separately for each viable subgroup.

General Research Design

Within Canada there is a vast potential of environmental factors which could influence test performance. Empirical research is necessary 1) to examine the potential effect of these factors upon selection utility, 2) to devise procedures for incorporating these new variables into the selection model and 3) to assess the feasibility and efficacy of moderator variable procedures in validating the Classification Battery - Men (CBM).

A study is currently in progress investigating seasonal, regional and language influences in the CBM. This research project will progress in three phases -

<u>Phase I</u>: is the examination of the 1971 enrolment population (men) to establish norms and to test the three hypotheses of seasonal, regional and language differences.
<u>Phase II</u>: involves the collection of criteria data for this population, first, to explore possible enhancement of CBM utility using season and region (initially) as moderator variables and second, to elucidate any selection bias for or against specific subgroups.
<u>Phase III</u>: is the evaluation and implementation of statistical procedures (e.g. moderated regression, different selection sequences for subgroups, etc.) to control factors demonstrated in Phase II.

Phase I is currently in progress and will be completed early in 1972. As criterion data accrues for the 1971 population, Phases II and III will be initiated. Although the initial purpose is to examine seasonal, regional and language influences in the CBM, this general schema may be subsequently employed to study other variables.

Phase I

In this preliminary investigation, season and region are viewed more as global indices of the numerous factors which could interact to produce regional or seasonal candidate differences. It is hoped that data from this study will guide the efficient selection of new variables to examine. In addition, it is not appropriate to analyze language as a moderator. The French CBM is essentially a direct translation of the English version but one cannot assume that these two versions measure precisely the same underlying psychological structure. Although corresponding aspects of each version may be examined, the perspective that one is dealing with two distinct batteries must be maintained. The most appropriate scheme to empirically compare English and French versions is through construct validation analyses where the factor structure underlying each battery is studied. Then, the two versions may be compared with respect to similarities and differences in the psychological structure tapped by each battery. It is hoped that this present project will furnish data towards this goal.

This project addresses three problems -

1. Are there seasonal fluctuations in the general quality of candidates available to the system depending upon the time of year?
2. Are there regional differences in candidates across Canada (e.g. East Coast vs Quebec vs Ontario vs Prairie Provinces vs West Coast)?
3. Do current testing procedures provide for optimal classification and assignment of English and French speaking applicants?

The reasoning underlying Hypothesis 1 evolves from the postulated influence of higher unemployment during the winter months attracting a different type of candidate than in other seasons. Whether this influence results in a higher,

lower or equal quality applicant on the average needs to be demonstrated empirically. One might particularly look for generally higher quality candidates applying at the end of the school year in late spring. Intuitively, regional differences in applicants, Hypothesis 2, seems a logical and fruitful research problem, especially when considering the more rural atmosphere present in the East Coast and Prairie Provinces in contrast to the predominately urban background of Ontario candidates, or when surveying the diffuse educational systems present across Canada. Since Hypothesis 3, English and French differences in test performance, has been demonstrated in the past (Skinner, Rampton and Keates, 1971) emphasis is upon providing data to upgrade current selection and classification standards.

Background Information

The CBM was developed to provide standardized ability measures of the recruit population (not officer candidates) and to facilitate optimal classification of applicants into trades. Seven tests (English and French versions) comprise the current CBM and each test is composed of objective multiple-choice or matching items. A description of each test follows -

1. GC3: a measure of general ability, used to differentiate applicants in terms of speed of thinking and capacity to learn (80 items, 30 min.).
2. CA2: a measure of clerical ability consisting of four sub-tests: spelling, grammar, punctuation and classification (86 items, 7.5 min).
3. AC2: a measure of speed and accuracy in arithmetic computations (175 items, 12 min.).
4. SP2: a test of general science (45 items, 20 min.).
5. MT2: a nonverbal test about tools and their uses (36 items, 10 min.).
6. EL2: a test of general electronics (50 items, 30 min.).
7. MK2: a test of basic mechanical and physical principles (37 items, 20 min.).

Candidates are tested throughout Canada at 15 Recruiting and Selection Units (RSU's). The age of these applicants generally averages 19.5 years with a standard deviation of 1.5 years. The GC3 is used as an initial screening device and only those applicants who pass the minimum cutoff proceed to the remaining tests in the battery. Complete test results are only available for those

candidates who are enrolled. Thus, one is dealing with a restricted sample compared to the overall Canadian population.

Reliability measures are available for 5 of these tests, Table 1 (from Skinner et al, 1971). Recent reliability estimates are not available for

---

Insert Table 1 about here

---

the two highly speeded tests, AC2 and CA2. Data from the previous CBM validation is reported by McInnes (1968). However, insufficient criterion data at that time curtailed a complete examination of English and French versions for all major trades.

Method

CBM results were collected for all enrolled applicants from the first six months of 1971. This data was analyzed according to -

A. seasonal vs regional factors within the English CBM.

B. seasonal differences within the French version.

An insufficient sample size impeded a seasonal vs regional analysis within the French CBM because the vast majority of French speaking Canadians reside in the province of Quebec.

In the first analysis, the factor season consisted of two levels: Quarter 1 (Jan-Mar) and Quarter 2 (Apr-Jun). Five levels of the regional factor were examined, ie.:

1. East Coast (provinces of Newfoundland, Nova Scotia, New Brunswick and Prince Edward Island).
2. Quebec.
3. Ontario
4. Prairie Provinces (Manitoba, Saskatchewan, and Alberta).

5. West Coast (British Columbia).

The first step involved a multivariate analysis of variance (MANOVA) for the two factor design employing a computer program, IBRMANOV, written by the University of North Carolina Psychometric Laboratory. Wilks' lambda criterion with an F approximation developed by Rao (1952) is used as the test of significance. This program also computes an univariate analysis of variance. Next, a stepwise discriminant analysis was conducted using the University of California Biomedical Computer Program, BMD07M.

However, because of resource constraints, the B analysis was conducted in the univariate case employing the simple two-tailed test of significance. When all data on the 1971 enrolment population is collected, multivariate procedures will also be used to expand this preliminary analysis.

Results and Discussion

A. Season vs Region (English CBM)

Total N for Q1 and Q2 was 1704. A list of Ns, means and standard deviations for each season by region cell is included in Table 2 and a test

---

Inserts Tables 2 and 3 about here

---

intercorrelation matrix is reproduced in Table 3. A cursory examination shows a general rise in test scores from east to west. In addition, many of the tests are highly correlated, for example, 0.795 between EL2 and SP2.

Results from the multivariate test of seasonal differences (S), regional differences (R) and season-region interaction (SR) are presented in Tables 4, 5, and 6 respectively. No significant seasonal effects (Hypothesis 1) were found.

Insert Tables 4, 5, and 6 about here

Similarly, the test of SR interaction proved non-significant. However, a significant regional effect (Hypothesis 2) was observed. Eigenvalues 1 through 4 and 2 through 4 were both significant at .001 level and roots 3 through 4 were significant at $P < .015$. Consequently, there are at least three orthogonal discriminant functions which distinguish the regional groups, Table 7.

Insert Table 7 about here

These three functions account for 98.5% of the total dispersion between groups, with the first function representing 68.5%, the second function 20.8% and the third function 9.2%. The corresponding canonical correlations between the seven CBM tests and the five geographical regions were 0.276, 0.154, and 0.106 respectively.

The largest contributers to group separation by region along the first discriminant function are the MT2 (tools and their uses) and GC3 (general mental ability). This function may be interpreted as differentiating the groups along a dimension of general, practical intelligence. Interpretation of the second and third discriminant functions, however, is more difficult and is of necessity tenuous in nature. The second discriminant function has high positive weighting on the SP2 (general science) and high negative weighting on the EL2 (electronics) and CA2 (clerical skills). Because of a high degree of common variance between the SP2 and EL2 (64%, see Table 3) and since the SP2 is weighted positively and the EL2 negatively, it must be the variance unique to each test which comprises the opposing poles of this dimension. The CA2 also contributes to the negative end of this dimension. Finally, the third discriminant function weights postively on the MK2 (mechanical and physical principles), MT2 and CA2, and negatively on the SP2.

In the stepwise discriminant analysis, one variable is entered at each

step into the set of discriminating variables. The variable entered has the largest F value, or, in other words, provides the greatest improvement in distinguishing the groups. The results, summarized in Table 8, demonstrate that the MT2 and GC3 were the first two variables to enter. This data corroborates results from the multiple discriminant analysis to identify the

---

Insert Table 8 about here

MT2 and CG3 as the largest contributors to regional variation.

To ascertain whether each of the five groups is distinct, an F matrix to test equality of means between each pair of groups is included, Table 9.

Insert Table 9 about here

Except for the East Coast and Quebec all regions were significantly different from one another. Consequently, four viable subgroups exist:

1. East Coast and Quebec
2. Ontario
3. Prairie Provinces
4. West Coast

To measure the extent of differentiation, Tatsuoka (1970) suggests a multivariate analogue to the estimated ω^2 as a total discriminatory power index. Substituting the appropriate values, it is found that about 11% of the variability in the discriminant space is attributable to regional differences. In other words, approximately 11% of the total variation of discriminant functions is accounted for by regional influences.

There are many possible reasons underlying this regional difference in candidates. Certainly, variations in educational systems, father's occupation,

urban-rural environment, etc, are potential factors. Perhaps a different "type" of individual is attracted to the Canadian Armed Forces in each locality because of regional differences in job opportunities. Alternatively, these observed subgroup differences in test performance may not reflect actual differences in ability and may be a function of tests that are biased for or against specific regional groups. At this point one may only speculate which factors are salient. Further, empirical investigation is mandatory to disentangle these potential influences.

In a study similar to the present, Guinn, Tupes and Alley (1970a) examined the influences of five demographic-cultural variables (race, educational level, geographic region, economic status, and city size) upon aptitude test scores in the US Air Force. They found significant relationships between the cultural variables combined and each aptitude test with the strongest relationships with tests measuring components of general intelligence. This finding is similar to this present investigation in that regional influences as represented by the first discriminant function differentiate groups along a dimension of measured, general and practical intelligence. In addition, they reported the weakest relationships with more situational specific tests (eg, psychomotor, memory) which are less dependent upon background experiences.

A second project (Guinn, Tupes and Alley 1970b) analyzed the fairness of selector aptitude indices to training criteria and found evidence of bias in 5 of 10 technical school groups. However, the type of bias varied, for example: area bias for Electrical Repair, race bias for Medical specialities, area and educational level bias for Communications Operations, and race, education and area bias for Engine Maintenance and Air Police. Focussing upon geographic area, they reported no common prediction error across all samples although individuals from the North-Northwest tended to be overpredicted and candidates from the Far West-Pacific Coast tended to be underpredicted.

B. Seasonal Differences (French CBM)

Results from the French version for Q1 and Q2 are presented in Table 10. Corresponding to the test for seasonal effects in the English CBM, no significant seasonal influences were observed in the French version.

Insert Table 10 about here

From a cursory examination of score profiles from both English and French versions, one interesting trend is that French candidates tend to score higher on the two clerical aptitude tests in relation to the remaining tests in their battery when compared with the English score profile. Implications and interpretation of this trend, however, must be mitigated by the inherent difficulties in comparing French and English batteries (see earlier discussion) and must await substantiation by further empirical evidence.

Summary

To elucidate possible seasonal and regional differences among applicants, CBM results from the first six months of 1971 were analyzed. No significant seasonal influences were found in either the English or French batteries, however a significant regional effect was present in the English version with a general rise in test performance (except for the clerical aptitude tests) from East to West across Canada. Four distinct subgroups were established: 1) Atlantic provinces and Quebec, 2) Ontario, 3) Prairie Provinces and 4) British Columbia. From a discriminant analysis it was found that the major discriminant function distinguished these groups along a dimension of general, practical ability. Possible reasons underlying this regional difference in candidates were discussed.

Phase I will be concluded early in 1972 when complete data for the 1971 enrolment population (men) is available. Phases II and III may then be initiated.

This project is but one component of the general program to assess and enhance utility of the CBM. The principal purpose of this specific component is to empirically evaluate a more individualized selection and classification strategy through a moderator variable approach. It is hoped that data from this study will elucidate the feasibility of this approach and guide the efficient selection of new variables to examine as potential moderators.

References

Ash, P. The implications of the Civil Rights Act of 1964 for psychological assessment in industry. American Psychologist, 1966, 21(8), 797-803.

Dunnette, M. D. Personnel selection and placement. Belmont, Calif.: Wadsworth, 1966.

Dunnette, M. D. A modified model for test validation and selection research. Journal of Applied Psychology, 1963, 47, 317-323.

Frederiksen, N., & Gilbert, A.C.F. Replication of a study of differential predictability. Educational and Psychological Measurement, 1960, 20, 759-767.

Frederiksen, N., & Melville, S.D. Differential predictability in the use of test scores. Educational and Psychological Measurement, 1954, 14, 647-656.

Ghiselli, E.E. Moderating effects and differential reliability and validity. Journal of Applied Psychology, 1963, 47, 81-86.

Ghiselli, E.E. Differentiation of individuals in terms of their predictability. Journal of Applied Psychology, 1956, 40, 374-377.

Guinn, N., Tupes, E.C., & Alley, W.E. Demographic differences in aptitude test performance. AFHRL-TR-70-15. Air Force Human Resources Laboratory, Personnel Research Division, May 1970. (a)

Guinn, N., Tupes, E.C., & Alley, W.E. Cultural subgroup differences in the relationships between Air Force aptitude composites and training criteria. AFHRL-TR-70-35, Air Force Human Resources Laboratory, Personnel Research Division, September 1970. (b)

Guion, R.M. Personnel selection. Annual Review of Psychology, 1967, 18, 191-216.

Guion, R.M. Employment tests and discriminatory hiring. Industrial Relations, 1966, 5, 20-37.

Hobert, R., & Dunnette, M.D. Development of moderator variables to enhance the prediction of managerial effectiveness. Journal of Applied Psychology, 1967, 51, 50-64.

Kirkpatrick, J.J., Ewen, R.B., Barrett, R.S., & Katzell, R.A. Testing and fair employment. New York: New York University Press, 1968.

Klein, S.P., Rock, D.A., & Evans, F.R. The use of multiple moderators in academic prediction. Journal of Educational Measurement, 1968, 5, 151-160.

McInnis, C.E. The development of a Common Selection and Classification Battery - Men for the Canadian Forces. Technical Report 68-8, CFPARU, Toronto, Canada: 1968.

O'Leary, B.S., Farr, J.L., & Bartlett, C.J. Ethnic group membership as a moderator of job performance. American Institutes for Research, Washington,DC, Technical Report Number 1. April 1970.

Rao, C.R. Advanced statistical methods in biometric research. New York: Wiley, 1952.

Rock, D.A., Barone, J.L., & Linn, R.L. A fortran computer program for a moderated stepwise prediction system. Educational and Psychological Measurement, 1967, 27, 709-713.

Rundquist, E.A. The prediction ceiling. Personnel Psychology, 1969, 22, 109-116.

Saunders, D.R. The "moderator variable" as a useful tool in prediction. Proceedings, 1954 invitational conference on testing problems. Princeton, N.J.: Educational Testing Service, 1955.

Skinner, H.A., Rampton, G.M., & Keates, W.E. Item analysis of the Canadian Armed Forces Classification Batteries. Research Note. CFPARU, Toronto, Canada: April 1971, in draft form.

Stricker, L.J. Compulsivity as a moderator variable. Journal of Applied Psychology, 1966, 50, 331-335.

Tatsuoka, M.M. Discriminant analysis: the study of group differences. Selected topics in advanced statistics, number 6, Champaign, Ill: Institute for Personality and Ability Testing, 1970.

Wolf, R. The measurement of environments. In A. Anastasi (Ed.), Testing problems in perspective. Washington D.C.: American Council on Education, 1966.

Zedeck, S. Identification of moderator variables by discriminant analysis in a multi-predictable group validation model. Journal of Applied Psychology, 1971, in press. (a)

Zedeck, S. Problems with the use of "moderator" variables. Psychological Bulletin, 1971, in press. (b)

Zedeck, S., Cranny, C.J., Vale, C.A., & Smith, P.C. Comparison of "joint" moderators in three prediction techniques. Journal of Applied Psychology, 1971, 55, 234-240.

Table 1

Reliability Estimates 1969 Sample

Test	N	Reliability KR-20
GC3 E	1475	0.92
GC3 F	635	0.86
MT2 E	1450	0.77
MT2 F	625	0.73
MK2 E	1379	0.68
MK2 F	632	0.62
SP2E	1462	0.79
SP2 F	641	0.75
EL2 E	1577	0.84
EL2F	633	0.82

Table 2

Descriptive Statistics English - CBM

Factor SR	N		GC3	CA2	AC2	SP2	MT2	EL2	MK2
11	221	M	45.09	37.74	64.86	22.28	18.65	24.88	24.01
		SD	7.95	9.83	28.47	5.31	4.85	7.52	4.37
12	45	M	44.40	38.58	65.36	21.67	18.18	24.47	23.22
		SD	8.33	9.00	21.15	5.77	4.69	7.34	4.17
13	368	M	46.84	37.47	64.16	23.91	20.32	25.45	24.30
		SD	8.71	11.66	28.91	7.75	5.82	9.00	5.98
14	288	M	48.00	39.70	66.66	25.50	20.94	27.69	24.50
		SD	8.87	9.37	26.57	6.22	5.30	7.72	5.19
15	142	M	48.85	39.67	62.20	24.86	22.13	27.52	26.45
		SD	8.27	9.28	25.43	6.53	5.13	8.09	5.10
21	114	M	45.07	37.07	68.83	23.34	18.01	25.39	23.46
		SD	7.83	8.89	27.46	5.70	4.66	7.63	4.26
22	25	M	45.36	42.28	71.72	21.84	18.80	24.60	24.72
		SD	9.50	10.83	28.48	5.23	4.15	6.83	4.36
23	269	M	46.22	36.14	62.27	23.88	20.66	25.09	24.58
		SD	9.22	11.53	28.70	7.18	5.58	8.25	5.37
24	158	M	48.23	38.90	64.96	25.25	20.14	27.86	25.08
		SD	8.98	11.99	29.01	7.60	5.74	9.11	6.47
25	74	M	49.22	39.68	62.31	25.96	22.13	28.23	26.38
		SD	8.66	8.72	28.29	5.19	4.79	7.06	4.43

Table 3

Test Intercorrelations

English Version	GC3	CA2	AC2	SP2	MT2	EL2
GC3						
CA2	0.50					
AC2	0.49	0.53				
SP2	0.43	0.46	0.29			
MT2	0.09	0.18	0.05	0.49		
EL2	0.51	0.45	0.38	0.80	0.48	
MK2	0.34	0.35	0.21	0.64	0.60	0.64

Table 4

Test of Season

Root	F	P less than	R Canonical
1	1.024	0.411	0.065

Table 5

Test of Region

Roots	F	P less than	R canonical
1 through 4	7.149	0.001	0.276
2 through 4	3.506	0.001	0.154
3 through 4	2.208	0.015	0.106
4	0.766	0.547	0.043

Table 6

Test of Season - Region Interaction

Roots	F	P less than	R canonical
1 through 4	0.892	0.628	0.090
2 through 4	0.632	0.870	0.061
3 through 4	0.501	0.891	0.049
4	0.249	0.911	0.024

TABLE 7

Standardized Discriminant Functions (Region)

	1	2	3
GC3E	0.780	0.107	0.072
CA2E	-0.132	-0.740	0.408
AC2E	-0.418	0.415	-0.189
SP2E	0.337	0.733	-1.120
MT2E	0.800	0.410	0.425
EL2E	-0.417	-1.019	-0.364
MK2E	-0.085	-0.444	0.572

TABLE 8

Stepwise Discriminant Analysis (Region)

Step #	Variable Entered	F Value to enter	Number of Variables	U Statistic
1	MT2E	19.3758	1	0.9564
2	GC3E	7.7536	2	0.9392
3	EL2E	7.5736	3	0.9227
4	AC2E	5.5982	4	0.9107
5	CA2E	3.7933	5	0.9026
6	SP2E	4.2308	6	0.8937
7	MK2E	1.9367	7	0.8896

F Matrix by Region

	East C.	Quebec	Ontario	Prairies
Quebec	1.1870			
Ontario	10.6008	5.8170		
Prairies	9.4585	5.3128	4.7649	
West C.	15.6529	7.7352	6.8284	4.2810

df 7,1693

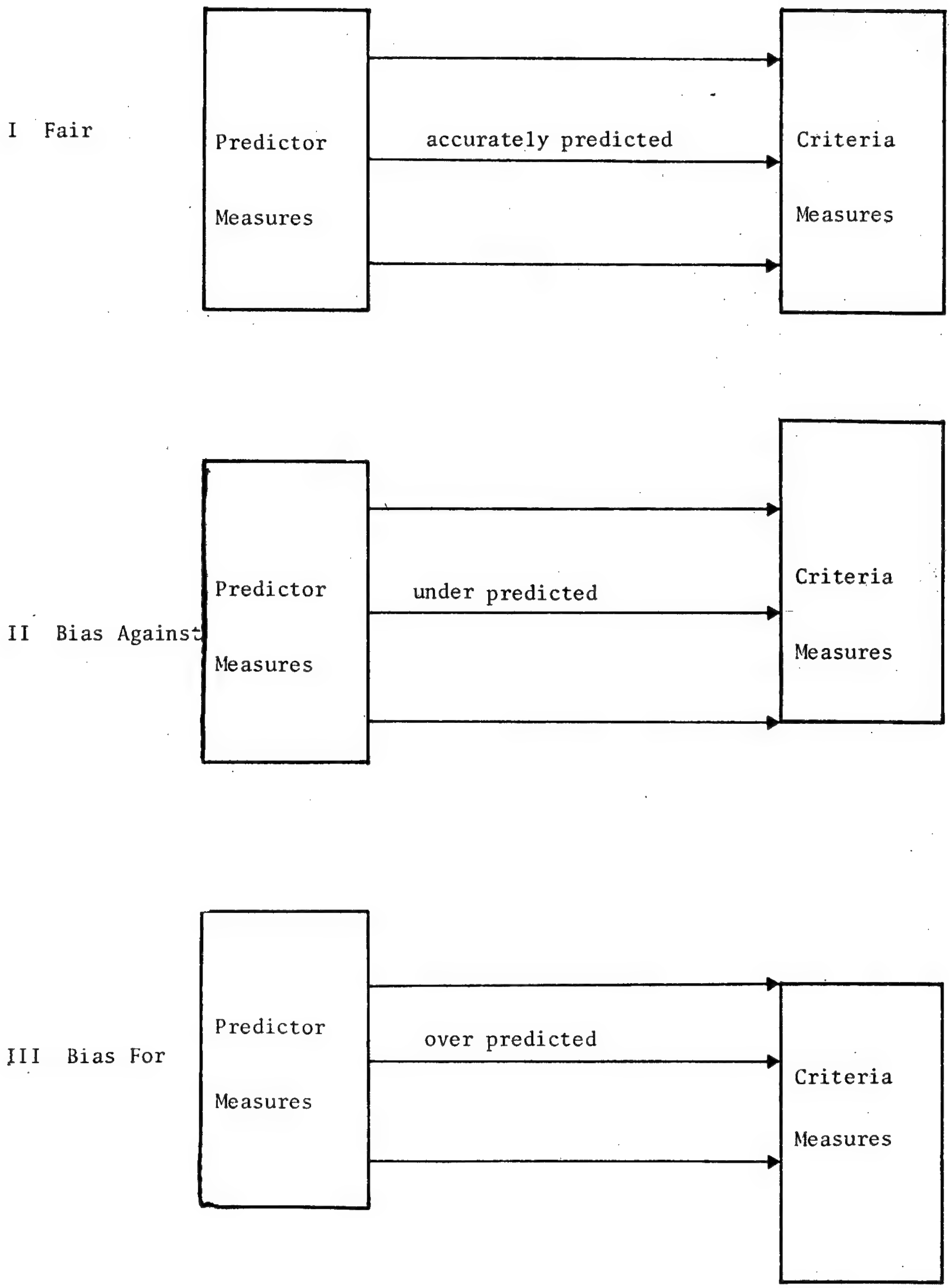
TABLE 10

Descriptive Statistics French CBM

	Quarter 1 N = 542		Quarter 2 N = 445	
	Mean	SD	Mean	SD
GC3F	42.44	8.04	42.59	7.98
CA2F	45.66	10.70	46.11	10.64
AC2F	74.67	26.27	74.40	23.90
SP2F	18.93	5.39	19.55	5.64
MT2F	17.95	4.73	18.48	4.87
EL2F	23.59	6.88	23.70	7.17
MK2F	22.63	4.11	22.96	4.22

FIGURE 1

SELECTION BIAS



Presentation

on

"PROPOSED BASIC CHANGES IN THE PERSONNEL STRUCTURE OF THE GERMAN ARMED FORCES"

by

Hermann O. Pfrengle,

German Observer Group, Aberdeen

in Cooperation with

LTC Ralf Rodenhauser,

Chief, German Observer Group, Aberdeen,

before the

13th Annual Conference of the

MILITARY TESTING ASSOCIATION

September 20 - September 25, 1971

Washington, D.C.

PROPOSED BASIC CHANGES IN THE PERSONNEL STRUCTURE OF THE GERMAN ARMED FORCES

Mr. Chairman, Ladies and Gentlemen:

On behalf of LTC Rodenhauser, Chief of the German Observer Group, Aberdeen, I would like to thank you for this opportunity to present a German contribution to this conference whose motto is "Improving Techniques and Procedures in Personnel Evaluation".

In keeping with the tradition of the presentations given at this conference, I will start out with an old joke which may, perhaps, characterize the situation we are in. In keeping with a Prussian tradition I would like to refer to the Prussian general who had his horse brought up for his daily morning ride. One morning the general mounted his horse face backwards, and, upon his adjutant's remark that he had mounted his horse face backwards, the general replied: "How do you know which way I want to go?!"

In breaking with Prussian tradition, I will briefly touch upon some entirely new proposed basic changes in the personnel structure of the German Armed Forces, as expressed in the "Study Report of the Personnel Structure Commission" of the German Ministry of Defense[1]. A decision on the implementation of this proposed concept has not been made by the German Government, as yet.

---

1. Personnel Structure Commission of the German Ministry of Defense, "The Personnel Structure of the Armed Forces" Study Report, June 1971 (German Title: Personalstrukturkommission des Bundesministers der Verteidigung: "Die Personalstruktur der Streitkräfte", Bericht der Personalstrukturkommission, Juni 1971)

The problem is: which way to go. The German Armed Forces, which, from the viewpoint of personnel structure, practically represent a mix of the draftee and volunteer principle, obviously require a new concept of personnel structure, more so in view of the growing trend toward systems integration which characterizes all highly industrialized societies. This basic restructuring appears necessary, on the one hand, in order to prevent an isolation of the military sphere from the other spheres of a highly mobile and pluralistic society, and, on the other hand, to assure combat effectiveness in the next decades which will be characterized by more intense technological progress than the present one.

In view of the contemplated change-over of the US Army to the volunteer principle it may be permitted to discuss here some aspects, on the basis of the aforementioned study, as they relate to the compatibility between military and civilian occupational criteria. This sociological problem which results from the relationship between the military and society in general, and from the relationship between a volunteer army and society in particular, has been stated by the American sociologist Robin Williams as follows: "The idea of a professional, permanent military establishment ... raises the specter of an insulated, powerful, authoritarian ... and politically ambitious social formation ... This consideration has been a major element in the tradition favoring the temporary citizen-soldier."[2] This American concept of the "citizen-soldier" found its German parallel in the "citizen in uniform".

The effort to arrive at a new concept of a German Armed Forces Personnel Structure is based on a directive mentioned

2. Robin M. Williams, "American Society", New York: Alfred A. Knopf, 1970, p. 269.

in the 1970 White Paper of the German Armed Forces.[3] A commission and concept management group was assigned the task "to develop, for the German Armed Forces, a new personnel concept"[4] which will enable the forces, in fully retaining their defense readiness, to adapt to new technological and societal developments over an extended period of time.

The commission's goals for this effort were determined by the following guidelines:

"The personnel structure of the Armed Forces results from the individual activities required to meet the mission of defense, and from their functional assignment to personnel utilization categories in a hierarchically structured organization which is to be managed in accordance with modern command and control methods.

This personnel structure provides for streamlining the forces such that they will be able to cope with all conceivable forms of war, as well as the utilization of new weapons systems.

The personnel structure provides the frame for determining career specialties and pay schedules. It determines training goals as well as type and extent of occupational promotion and advancement of military personnel.

It further provides important presuppositions for optimal integration of the forces into the society."[5] In redesigning

---

3. The Federal Minister of Defense, "White Paper 1970 on the Security of the Federal Republic of Germany and on the State of the German Federal Armed Forces", May 1970.

4. "The Personnel Structure of the Armed Forces", p. 8.

5. Ibid., p. 9.

the personnel structure - especially under volunteer army aspects - the forces must take into account the given labor market conditions by aptitude and skill selection and promotion, by providing vertical occupational mobility, as well as by utilizing a performance evaluation and rating system which properly accounts for occupational demands placed upon the individual. The principle criteria for such personnel measures are personnel required and adequacy of costs versus net efficiency, i.e., cost effectiveness.

The interrelationships of personnel utilization categories, areas of responsibility, military ranks and pay scales indicate that the current personnel management system of the German Armed Forces is no longer adequate to meet the requirements for the next decades. The commission describes the present personnel situation as follows: "The relatively great differences, for example in the degrees of occupational demands placed upon individuals performing the same activities, are the cause of a considerable degree of job dissatisfaction. These differences are not taken into account by the current personnel evaluation, rating and pay system, and, therefore, have unfavorable effects on the personnel. At the present time, a career advancement, in terms of higher pay, is possible only via a higher military rank. The consequence is an accumulative concentration of relatively high military ranks, without the corresponding expansion of the base of the "take-off" ranks.[6]

From this follows that the ever-increasing number of individual activities requires not only their systematic reassessment. They must also be ordered into personnel utilization categories and must be assigned to areas of responsibility.

6. Ibid., p. 51.

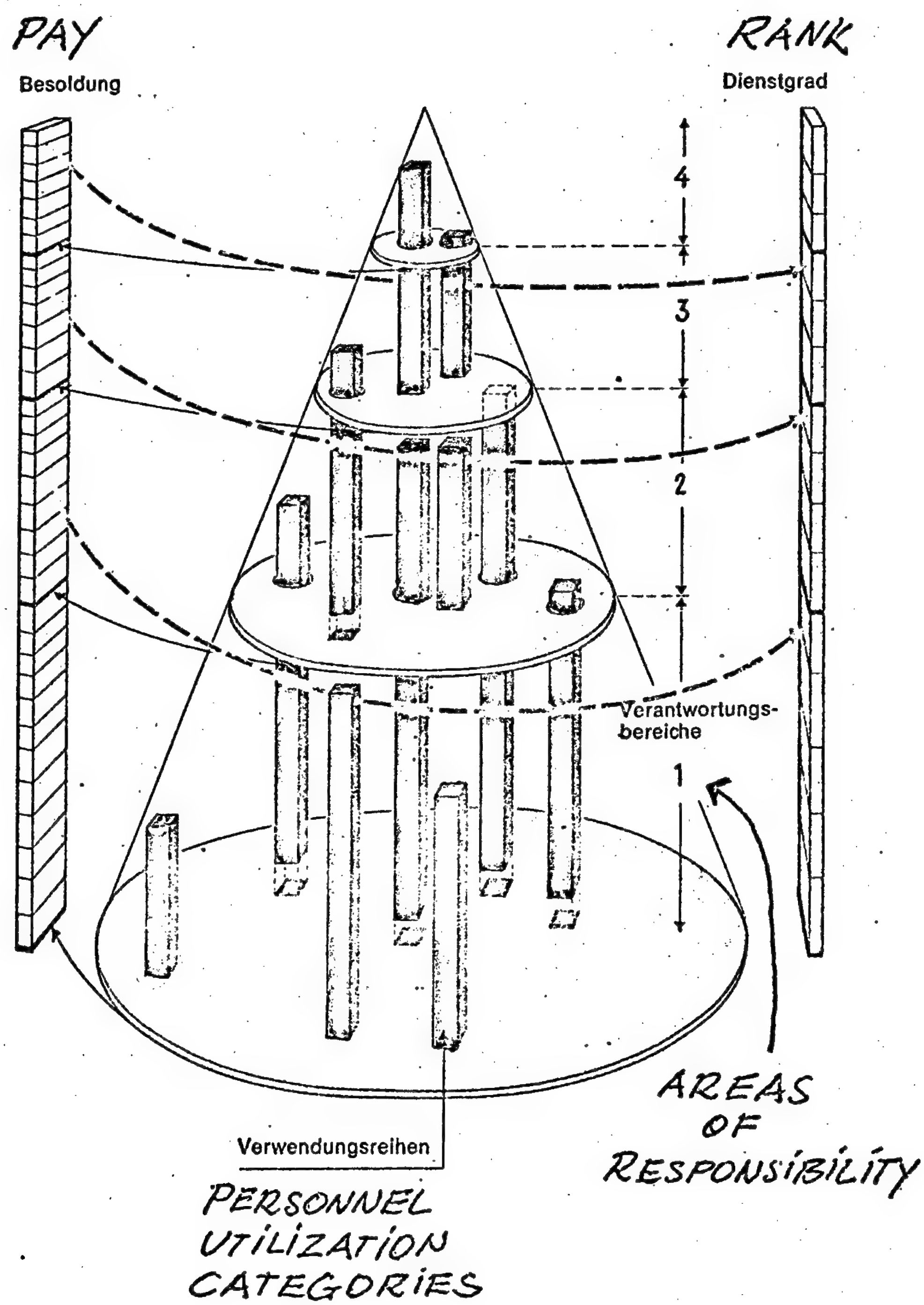
The commission comes to the conclusion that individual advancement, and along with it, the pay, should no longer depend on the military rank." If a soldier meets the requirements for another activity in the same personnel utilization category, which is assessed higher by the job evaluation system, then this soldier should receive a higher pay in that activity. If this activity lies within the same area of responsibility, then his military rank will remain the same - unless the requirements for receiving a promotion are met in correspondence with the graduated degrees of responsibility within that area of responsibility. If a personnel utilization category extends across several areas of responsibility, and if the new activity of the soldier falls into a higher area of responsibility, then he will be promoted to that rank which corresponds with his new, higher area of activities."[7]

The structural and functional interrelationships between personnel utilization category, area of responsibility, military rank and pay are depicted by the following model*:

7. Ibid., p. 53

* Based on questions raised following the presentation, this model will be explained in the Appendix.

MODEL OF INTERRELATIONSHIPS BETWEEN PERSONNEL UTILIZATION CATEGORIES, AREAS OF RESPONSIBILITY, RANKS AND PAY



(Source: "The Personnel Structure of the Armed Forces", p. 59)

Some of the major milestones for the step-by-step realization of this new personnel structure, by earliest possible dates, are as follows:

Milestone Description	Earliest possible Completion Date
Start listing of all existing training contents and characteristics	1 July 71
Start setting-up the management organization	2 Aug 71
Start analysis and selection of typical activities	1 Apr 72
Completion of individual skill demand profiles	1 Feb 75
Completion of personnel utilization categories	1 Feb 75
Criteria for ranks, areas of responsibility and degrees of responsibility completed ..	1 Apr 75
Start establishing coordinated personnel data collection and processing centers ...	1 Oct 75
Development of an integrated personnel structure information system completed ...	1 Oct 77
Enactment of new Military Service Legislation	1 Sep 78
Determination of overall costs of the personnel structure concept completed	2 Jan 79
New personnel structure concept functional	1 Mar 81

In conclusion, the German Armed Forces are faced with a ten-year effort of structural and functional integration of the individual military and civilian into their organization, starting from the overall defense task, and breaking it down into subtasks, and, finally, into individual positions. The desired goal is the structural and functional integration of the individual soldier, government official and

employee into the defense organization. This effort is expected to result in an optimal adaptation of the military sphere to the given conditions of an increasingly mobile industrial society.

APPENDIX

Description of Model, Page 6

The model reflects the activity-oriented assessment and pay, and their interrelationships with responsibility and rank. The study report states that "... the job assessment and pay system must be oriented along the lines of the individual activity. The prerequisite for such assessment and pay, as shown, is a system of activity description, analysis and evaluation which covers the dimension of the activities in a differentiated manner. Furthermore, this system provides the basis for the design of personnel utilization categories, individual demand profiles, etc."[8]

Primary Elements of the Model

The Ranks

The clearcut command, control and execution of the military mission requires the retention of the ranks. The rank is the visible symbol of authority. Within each area of responsibility there will be graduations of responsibility whose visible symbol will be the rank.

The Areas of Responsibility

Findings derived from the present personnel structure permit a general categorization of responsibilities into four areas. However, the precise determination of the number of areas of responsibility will not be possible until after the descriptions, analyses, assessments and evaluations of all individual activities, and their incorporation into categories of personnel utilization, have been completed.

The study report envisions four areas of responsibility:[9]

---

8. "The Personnel Structure of the Armed Forces", p. 53.
9. Ibid., p. 55 ff.

Area of Responsibility No. 1

Simple action responsibility (and most simple command and control responsibility, as required);

- man on guard duty,
- man at his equipment,
- man in a simple technical activity, etc.

Requirements for performing activity:

Knowledge and skill for performing activities of a simple nature, under supervision or guidance.

Training goal comparable to the following levels: skilled laborer, "Geselle"*, foreman, crew leader, assistant commercial clerk, or ability to perform corresponding administrative activity.

Area of Responsibility No. 2

Action responsibility; simple command and control responsibility;

- guard responsibility; in charge of guard detail;
- technical responsibility for equipment components; responsibility for technical sub-units;
- tactical command and control of small military units (squad/platoon).

Requirements for performing activity:

Profound knowledge and abilities to perform assigned missions, possibly under guidance.

Training goal comparable to the following levels: tactical training enabling incumbent to command and control small military units; technical military training, chief foreman, technician, ability to perform corresponding administrative activity.

Area of Responsibility No. 3

Command and control responsibility; upper-level action responsibility;

- combined responsibility for personnel and equipment;

* The German "Geselle" is roughly equivalent to "journeyman"; he has to pass a final examination after 3 years of trade-school.

- superior with court-martial jurisdiction;
- command and control of units and task forces (tactically/operationally);
- preparing requirements; managing and evaluating weapons systems and technical installations;
- executing missions without supervision or guidance, based on profound knowledge and fully developed abilities.

Training level: completed university education (master's).
Training goal: general staff officer, and/or activity on the intermediate management level.

Area of Responsibility No. 4

Upper level command and control responsibility (highest level action responsibility, as required);

- ability to make judgements and decisions in the area of defense policies;
- profound knowledge, primarily in questions dealing with military policies, defense economy, political and educational aspects of societal policies;
- command and control of forces in the strategic/operative range.

Training level: completed Defense College (General/Admiral Staff).

Pay

These categories of personnel utilization, shown as red columns in the model, comprise mutually comparable individual activities. The basically new and important point is the fact that the individual levels separating the areas of responsibility have become penetrable, as shown in the model. This had been possible under contemporary personnel

structure in isolated cases only

Summary of Model Description

The model illustrates the commission's proposal which is centered around the personnel utilization category and its direct relation to the area of responsibility, as the nuclear element of a soldier's career. Pay and rank are functions of it. Arrangement and interrelationships between these four elements provide for an increase in vertical mobility which, in accordance with the commission's concept, will take into account the increased mobility of the entire society.

BY-LAWS OF THE
MILITARY TESTING ASSOCIATION

Article I - Name

The name of this organization shall be the Military Testing Association.

Article II - Purpose

The purpose of this association shall be to:

A. Assemble representatives of the various armed services of the United States, Canada, Royal Australian Air Force, and such other nations as might be admitted subsequently to the existing association to discuss and exchange ideas concerning military testing or job proficiency evaluation of enlisted personnel.

B. Review, study, and discuss the mission, organization, operation, and research activities of the various associated armed services agencies engaged in the testing of military personnel.

C. Foster improved evaluation through exploration of new techniques and presentation of improved procedures for test development, processing of examination returns, evaluation of test results through objective and vigorous statistical analysis, and subjective discernment of item and examinee behavior.

D. Promote among the agencies of the various associated armed services engaged in military testing or job proficiency evaluation cooperation in the exchange of testing material, items, and procedures where appropriate.

E. Stimulate research in the area of testing.

F. Promote the evaluation of military personnel as a scientific adjunct to modern military personnel management.

Article III - Membership

The following shall be full members of the Military Testing Association:

A. All active duty military and civilian personnel permanently assigned to an agency of the associated armed services having primary responsibility for the conduct of military testing or job proficiency evaluation in the respective services.

B. All military or civilian personnel assigned to an agency of the associated armed services having secondary responsibility for the conduct of job proficiency evaluation who are directly responsible for any phase of military testing or job proficiency evaluation.

C. All military personnel temporarily assigned to an agency of the associated armed services having primary responsibility for the conduct of job proficiency evaluation and who are actively engaged in the program.

D. All civilian and active duty military officer personnel permanently assigned to a Directorate, office, bureau, or secretartiate exercising direct command over associated armed services agencies holding primary responsibility for the conduct of military testing or job proficiency evaluation.

E. Any individual, military or civilian, designated by the Steering Committee of the association, who is engaged in any phase of testing in the associated military establishment, government, business, industry, or educational institution.

Article IV - Dues

No annual dues shall be levied against the members, full or associate, of the association.

Article V - Officers

A. The officers of this association shall consist of a President and a Secretary.

B. The President of the association shall be the Commanding Officer of the associated armed services agency hosting the annual conference of the association. The term of office of the President of the association shall begin at the close of the annual conference of the association and shall expire at the close of the immediately following annual conference of the association.

C. It shall be the duty of the President to organize and host the annual conference of the association held during his term of office, to preside at meetings, to act as ex officio chairman of the Steering Committee of the association, and to perform the customary duties of a president.

D. The Secretary of the association shall be filled through appointment by the President of the association. The term of office of the Secretary shall be the same as that of the President.

E. It shall be the duty of the Secretary of the association to keep the records of the association, and to issue notices for conferences. The Secretary shall also perform such additional duties and take such additional responsibilities as the President may delegate to him.

Article VI - Steering Committee

A. There shall be a Steering Committee of the Association that shall consist of permanent members as follows:

1. Permanent members shall be the Commanding Officers of the member armed services agencies exercising primary responsibility for the conduct of military testing or job proficiency evaluation in the respective services.

2. In addition to the regular permanent military member as stated above, each of the associated armed services agencies exercising primary responsibility for the conduct of military testing or job proficiency evaluation in the respective services shall have their ranking civilian professional employee and no more than one additional appointee as permanent members.

B. Permanent members of the Steering Committee shall serve as members of said committee during their tenure.

C. The President of the association shall be Chairman of the Steering Committee of the association and the term of the office shall be one year. The chairmanship can also be delegated by the President or if he so desires a chairman may be selected through election.

D. A secretary of the Steering Committee shall be appointed by the Chairman of the Steering Committee. The term of office of the Secretary of the committee shall be one year.

E. The Steering Committee shall have general supervision over the affairs of the association and shall have responsibility for all activities of the association. The Steering Committee shall conduct the business of the association in the interim between annual conferences of the association.

F. Meetings of the Steering Committee shall be held during the annual conferences of the association and at such other times as requested by the President of the association or the chairman of the Steering Committee. A majority of the permanent and appointed members of the Steering Committee shall constitute a quorum.

Article VII - Meetings

A. The association shall hold a conference annually.

B. The annual conference of the association shall be hosted by the agencies of the associated armed services exercising the primary responsibility for military testing or job proficiency evaluation in order of the following rotating schedule:

United States Army
United States Navy
Canadian Armed Forces
United States Air Force
United States Coast Guard
United States Marine Corps

C. The annual conference of the association shall be held at a time and place determined by the host agency. The membership of the association shall be informed at the annual conference of the place at which the following annual conference will be held. The host agency shall inform the Steering Committee of the time of the annual conference not less than 6 months prior to the conference.

D. The host agency shall exercise planning and supervision over the program of the annual conference.

Article VIII - Committees

A. Standing committees may be named from time to time, as required, by vote of the membership of the association. The chairman of each standing committee shall be appointed by the President of the association with the counsel and approval of the Steering Committee. Members of standing committees shall be appointed by the President of the association in consultation with the chairman of the committee in question. Chairmen and committee members shall serve in their appointed capacities at the discretion of the President. The President shall be an ex officio member of all standing committees.

B. The term of office of committee chairmen and committee members shall continue no longer than the term of office of the President who made the appointment.

C. The President with the counsel and approval of the Steering Committee shall appoint such ad hoc committees as are needed from time to time. An ad hoc committee shall serve until its assigned task is completed or for the length of time specified by the President in consultation with the Steering Committee.

D. All ad hoc committees shall be dissolved when the term of office of the President who appointed them expires, with the

exception that the incoming President and Steering Committee may continue such appointments to ad hoc committees beyond this date.

E. All standing committees and all ad hoc committees shall clear their general plans of action and new policies through the Steering Committee, and no committee or committee chairman shall enter into relationship or activities with persons or groups outside of the association that extend beyond the approved general plan of work without the specific authorization of the Steering Committee.

Article IX - Amendments

A. Amendments of these By-Laws may be made at any annual conference of the association.

B. Amendments of these By-Laws may be made by majority vote of the assembled membership of the association provided that the proposed amendments shall have been approved by a majority vote of the Steering Committee.

C. Proposed amendments not approved by a majority vote of the Steering Committee shall require a 2/3rds vote of the assembled membership of the association.

Article X - Voting

All members in attendance shall be voting members.

Article XI - Enactment

This constitution shall be in force immediately on its acceptance by a majority of the assembled membership of the association and or amended (in force 17 September 1970).

Report of the 1971 Steering Committee Meeting

The 1971 Military Testing Association (MTA) Steering Committee convened at the Statler Hilton Hotel, Washington, D.C., at 1330, 20 September 1971. The following members were present:

U. S. ARMY

Dr. R. O. Waldkoetter
Mr. K. C. Liebfried

U. S. MARINE CORPS

Col L. T. Erickson, Chairman
Mr. E. A. Dover
LtCol M. E. Bane, Secretary

U. S. NAVY

Capt C. E. McMullen
Mr. C. S. Winiewicz
Mr. C. J. Macaluso

U. S. AIR FORCE

Col O. A. Berthold
Col R. S. Hoggatt

U. S. COAST GUARD

Cdr K. R. Depperman
Mr. J. A. Burt

CANADIAN ARMED FORCES

LtCol R. K. Acheson
Cdr W. H. Northey

ROYAL AUSTRALIAN AIR FORCE

Sqn Ldr J. W. K. Fugill

The following items and actions were considered or acted upon by the Steering Committee:

1. It was confirmed that the 1972 and 1973 conferences will be hosted by the U.S. Navy and the U.S. Air Force, respectively. The 1972 conference will convene in Lake Geneva, Wisconsin; the 1973 conference, in San Antonio, Texas. The hosting sequence beyond 1973 has not yet been agreed upon.

2. In view of the manner in which the association is growing, it was agreed that a reappraisal of its initial objectives was required. A study of Article II of the By-Laws has been undertaken.

3. The Canadian Armed Forces requested by letter that they be allowed to limit their participation in the association to that of observers at the annual conference, to include the presentation of invited papers. The action recommended by the Steering Committee, in keeping with the flexible membership policy of the MTA, was that the Canadian membership is suspended but not terminated. It was further recommended that a letter be directed by the incoming MTA president to the Canadian military authorities to reaffirm the association's interest in the CAF participation at whatever resource level they may wish to invest.

4. The Steering Committee voted favorably on the application for membership of the Royal Australian Navy.

ATTENDEES
13TH ANNUAL CONFERENCE
MILITARY TESTING ASSOCIATION

LtCol R. K. Acheson	Directorate of Training Canadian Forces Headquarters Ottawa 5, Ontario Canada
Mr. Douglas Adair	USCG TRACEN Systems Section Governors Island New York, N. Y. 10004
Mr. Antoine Al-Haik	Defense Language Institute Systems Development Agency Presidio of Monterey, Calif. 93940
Mr. Robert E. Anneser	U. S. Postal Service 8120 Woodmont Ave., Room 531 Bethesda, Maryland 20014
CWO-2 F. J. Appl, Jr.	Enlistment Procurement Headquarters Marine Corps Washington, D. C. 20380
LtCol M. E. Bane	Headquarters Marine Corps (Code A01B) Washington, D. C. 20380
Col J. M. Bannan	Headquarters Marine Corps DC/S (Air), AAZ Washington, D. C. 20380
Mrs. Ruth Barefield	U. S. Army Aviation School Ft. Rucker, Ala. 36360
Col A. E. Bench	Head, Training & Education Branch Headquarters Marine Corps G-3 Division Washington, D. C. 20380
Mr. Louis S. Berger	Southwest Research Institute P. O. Drawer 28510 San Antonio, Texas 78228
Col O. A. Berthold	Vice Commander Air Force Human Resources Lab. Brooks Air Force Base, Texas 78235
2ndLt Roger M. Beverage	The Judge Advocate General's School Charlottesville, Va. 22901

Cdr Alfredo Beyer R.	Peruvian Navy Ministerio De Marina Lima, Peru, S. A.
Mr. John Bondaruck	National Security Agency Department of Defense Ft. Meade, Md. 20730
Dr. J. W. Bowles	Air Force Human Resources Lab Lackland Air Force Base, Texas 78236
Mr. John Brand	USAEEC Ft. Harrison, Indiana 46249
Dr. Clay V. Brittain	USAFI 2318 S. Park Street Madison, Wis. 53713
Mr. Frederick C. Bruner	Data-Design Laboratories 801 N. Pitt Street Alexandria, Va. 22314
Mr. John A. Burt	U. S. Coast Guard Institute P. O. Substation 18 Oklahoma City, Okla. 73169
Col G. Caridakis	Headquarters Marine Corps (Code A01H) Washington, D. C. 20380
LtJG David Chittenden	U. S. Coast Guard Institute P. O. Substation 18 Oklahoma City, Okla. 73169
CWO2 Joseph T. Cook	U. S. Coast Guard Training Center Governors Island New York, N. Y. 10004
GySgt David A. DeOre	USMC Communication-Electronics School 29 Palms, California 92278
Cdr K. R. Depperman	U. S. Coast Guard Institute P. O. Substation 18 Oklahoma City, Okla. 73169
Capt Robert E. Dorgan	USAAVNS Attn: Anal Div. DOI Fort Rucker, Ala. 36360

Mr. E. A. Dover	Headquarters Marine Corps (Code A01B) Washington, D. C. 20380
1stLt Kirt E. Duffy	Air Force Human Resources Lab Lackland Air Force Base, Texas 78236
Mr. Erling A. Dukerschein	U. S. Naval Examining Center Great Lakes, Illinois 60088
Lt W. F. Ellis	Special Projects Office (Code 1542) Department of the Navy Washington, D. C. 20390
Col L. T. Erickson	Headquarters Marine Corps (Code A01B) Washington, D. C. 20380
1stLt Raymond R. Erickson	USA Adjutant General School Administrative Schools Center Ft. Benjamin Harrison, Ind. 46220
Capt A. R. Finlayson	Headquarters Marine Corps (Code DFA) Washington, D. C. 20380
Mr. Paul P. Foley	Naval Personnel Research and Development Laboratory Washington, D. C. 20390
Capt Bruce A. Fournier	Personnel Applied Research Unit 1107 Avenue Road Toronto 305, Ontario, Canada
Mr. Sidney Friedman	Office of the Assistant Secretary of the Navy (M&RA) Washington, D. C. 20350
SqnLdr John Fugill	Royal Australian Air Force Exchange Officer, AFHRL Lackland Air Force Base, Texas 78236
Mr. A. Gardner	Senior Psychologist (Naval Division) Ministry of Defence (Navy) London SW1, England
LCdr Dalton L. Garrison	U. S. Coast Guard Headquarters 400 7th Street, S. W. Washington, D. C. 20024

Mr. Vincent A. Geraldi	USA Signal School ATSSC-DOI-EVD Fort Monmouth, N. J. 07703
Dr. Arthur C. F. Gilbert	Naval Personnel Research and Development Laboratory Bldg 200, Washington Navy Yard Washington, D. C. 20390
LtCol E. J. Godfrey	Training Evaluation and Support Section Training and Education Branch G-3 Division, HQMC Washington, D. C. 20380
Mr. H. Wm. Greenup	Education Center MCDEC Quantico, Virginia 22134
Capt Harry H. Greer, Jr.	5 Kingwood Drive Poughkeepsie, N. Y. 12601
Major H. G. Halliday	Canadian Armed Forces School of Instructional Technique (CFSIT) CFB Borden Borden, Ontario, Canada
Mr. Billy H. Hannaford	Data-Design Laboratories 801 N. Pitt Street Alexandria, Va. 22314
Mr. Lawrence W. Head, Jr.	USAF Security Service School Goodfellow Air Force Base, Texas 76901
Maj Harold E. Hemming	Canadian Forces Headquarters CG/DGOM/DMPC4 Job Analysis Section Mortimer Bldg 201 Ottawa, Ontario, Canada
Mr. Arthur G. Hermansen	Enlisted Evaluation Center U. S. Army Ft. Benjamin Harrison, Ind. 46226
Maj Charles D. Herrera	Chief, Evaluations USA Military Police School Ft. Gordon, Ga. 30905

Col R. S. Hoggatt	Air Force Human Resources Lab Personnel Research Division Lackland Air Force Base, Texas 78236
Mr. Jack E. Hohreiter	USAEEC Ft. Benjamin Harrison, Ind. 46249
Dr. John J. Holden	USA Ordnance Center & School Aberdeen Proving Ground, Md. 21005
Col D. J. Hunter	Head, Human Relations Branch G-3 Division, HQMC Washington, D. C. 20380
Mr. B. W. Kane	U. S. Coast Guard Training Center Governors Island New York, N. Y. 10004
Capt Wayne E. Keates	Canadian Forces Personnel Applied Research Unit 1107 Avenue Road Toronto 305, Ontario, Canada
Miss Anna May Kelleher	U. S. Coast Guard Training Center Governors Island New York, N. Y. 10004
Capt W. B. Klages	Headquarters Marine Corps (Code ABR) Washington, D. C. 20380
Mr. Richard S. Kneisel	USA Infantry School Ft. Benning, Ga. 31905
Maj G. A. Knudson	Headquarters Marine Corps G-3 Division Washington, D. C. 20380
Capt Fred W. Koehler	Headquarters Marine Corps Division of Reserve Washington, D. C. 20380
Mr. Delbert E. Kohl	Marine Corps Institute Marine Barracks, Box 1775 Washington, D. C. 20013
Dr. J. B. Koplyay	Air Force Human Resources Lab Lackland Air Force Base, Texas 78236

LtCol Bryce R. Kramer	USA Infantry School Ft. Benning, Ga. 31905
Mr. Melvin W. Lackey	Training Publications Division, NPPSA Bldg 220, WNY Washington, D. C. 20390
Mr. Richard Lanterman	Headquarters, U. S. Coast Guard 400 7th Street, S. W. Washington, D. C. 20591
Maj R. L. Lary	Headquarters Marine Corps Washington, D. C. 20380
Mr. Kenneth C. Liebfried	USAEEC Ft. Benjamin Harrison, Ind. 46226
LtCol R. E. Loehe	S and R Division Development Center MCDEC Quantico, Va. 22134
LtCol Theodore J. Lutz, Jr.	Headquarters Marine Corps DC/S (Air) Washington, D. C. 20380
Mr. Charles J. Macaluso	U. S. Navy Examining Center Bldg 2711 Great Lakes, Illinois 60080
Mr. Herman A. Mahnen	Enlisted Evaluation Center U. S. Army Ft. Benjamin Harrison, Ind. 46226
Mr. E. F. Magdarz	IBM Corp Department 915 Poughkeepsie, N. Y. 12602
Dr. Milton H. Maier	USA Behavior and Systems Research Lab 1300 Wilson Blvd. Arlington, Va. 22209
Col J. W. Marsh	Headquarters Marine Corps (Code AO1M) Washington, D. C. 20380
Maj David E. Marz	AIAOS Maxwell Air Force Base, Ala. 36112

Capt C. E. McMullen	U. S. Navy Examining Center Bldg 2711, NTC Great Lakes, Illinois 60088
Maj D. F. Mead	Headquarters, ATC, XPTT Randolph Air Force Base, Texas 78148
Maj James K. Miller	Headquarters Marine Corps G-3 Division Washington, D. C. 20380
Mr. Tom E. Miller	Naval Aviation Schools Command' Naval Air Station Pensacola, Fla. 32504
Mr. John W. Murphy	Finance School (USAFS) Ft. Benjamin Harrison, Ind. 46216
Mr. Johnny Nelson	USA Missile and Munitions Center and School Redstone Arsenal, Ala. 35809
Cdr W. H. Northey	Director of Personnel Applied Research - CEHQ Ottawa, Ontario, Canada
LtCol Richard G. Oakes	Division of Veterinary Medicine Walter Reed Army Institute of Research Walter Reed Army Medical Center Washington, D. C. 20012
Dr. Virgil J. O'Connor	Education Testing Service 960 Grove Street Evanston, Illinois 60201
Maj George Oetting	Academic Instructor & Allied Officer School Maxwell Air Force Base, Ala. 36112
GySgt David E. Oliver	Headquarters Marine Corps (Code DFM) Washington, D. C. 20380
Dr. Adolf Panitz	National Occupational Competency Testing Project Rutgers University, N. J. 08903
Mr. Gordon Parrish	U. S. Coast Guard Training Center Systems Section Governors Island, N. Y. 10004

Miss Ruth O. Peters	U. S. Postal Service 8120 Woodmont Avenue, Rm. 531 Bethesda, Maryland 20014
Mr. Hermann O. Pfrengle	FRG OBS GRP HQ TECOM APG Aberdeen Proving Ground, Md. 21005
Mr. George V. Porter, Jr.	USAF Academy Colorado Springs, Colorado 80840
SqnLdr B. N. Purry	Royal Air Force Headquarters Training Command Brampton, Huntington, England
LtCol Ralf Rodenhauser	HQ USATECOM Attn: AMSTE - FRG Aberdeen Proving Ground, Mr. 21005
Mr. Hal H. Rosen	Personnel Research Division Bureau of Naval Personnel Washington, D. C. 20370
Mr. William A. Sands	Naval Personnel Research and Development Laboratory Washington Navy Yard Washington, D. C. 20390
LCdr Ken G. Schultz	Naval Attache, Australian Embassy 1601 Mass Avenue, N. W. Washington, D. C. 20036
Mrs. Genevieve K. Schutter	U. S. Postal Service 8120 Woodmont Building Bethesda, Md. 20014
LtCol H. L. Scott, III	Headquarters Marine Corps (Code AO1H) Washington, D. C. 20380
Capt F. H. Schwarz	Headquarters Marine Corps DC/S (Air), AAZ Washington, D. C. 20380
Dr. E. H. Shuford	SM Corporation 406 Citizens First National Bank Building Tyler, Texas 75701

Mr. H. A. Skinner	Personnel Applied Research Unit 1107 Avenue Road Toronto 305, Ontario, Canada
Mr. Mabron H. Smith	USA Missile & Munitions Center Redstone Arsenal, Ala. 35809
Capt F. G. Snocker	Headquarters Marine Corps (Code A01B) Washington, D. C. 20380
Mr. E. P. Stichman	Bureau of Naval Personnel Washington, D. C. 20370
Mr. D. J. Sullivan	USA Security Agency Training Center and School USASATC&S Ft. Devens, Mass. 01433
Mr. Terry Y. Takahashi	Defense Intelligence School USNS Anacostia Annex Washington, D. C. 20390
Mr. Harry W. Vinicombe	U. S. Postal Service 8120 Woodmont Avenue, Rm. 531 Bethesda, Md. 20014
Mr. Raymond O. Waldkoetter	USA Enlisted Center & School Ft. Benjamin Harrison, Ind. 46249
Mr. Frank B. Walsh	Commanding Officer USA Security Agency Ft. Devens, Mass. 01433
Miss Martha E. Weaver	USA Chemical Center & School Ft. McClellan, Ala. 36201
Mr. L. A. Wedell	Director, Naval Education and Training Office of CNO 801 N. Randolph Arlington, Va. 22203
MGen E. B. Wheeler	Headquarters Marine Corps (Code A01) Washington, D. C. 20380
Mr. Gene T. Whitney	HQ USASESS ODDLP MOS Testing Dev. Ft. Gordon, Ga. 30905

Maj E. M. Wieler	Headquarters Marine Corps (Code A01C) Washington, D. C. 20380
Maj Robert E. Wilkinson	Personnel Research Division Air Force Human Resources Lab Lackland Air Force Base, Texas 78236
Mr. Harold A. Williams	USA Ordnance Center & School 529 Windemere Drive Aberdeen, Md. 21001
Mr. Walter S. Williams	Evaluation Division USA SE Signal School Ft. Gordon, Ga. 20905
Mr. Richard C. Willing	USCG - MVP 400 7th Street, S. W. Washington, D. C. 20590
Mr. C. S. Winiewicz	U. S. Naval Examining Center Bldg 2711, NTC Great Lakes, Illinois 60088
Lt Richard W. Wright	U. S. Coast Guard Merchant Vessel Personnel 400 7th Street, S. W. Washington, D. C. 20590
Mr. Ted Yellen	Naval Personnel Research and Development Laboratory Washington Navy Yard Washington, D. C. 20390
Mr. William J. York, Jr.	USA Military Police School Ft. Gordon, Ga. 30905